



# Программа Геном Человека



- Секвенирование генома человека практически полностью закончено.
- Информация о последовательностях человека в базе данных GenBank растет экспоненциально.
- База данных dbSNP содержит **4 268 472** записи об ОНП (на 16 мая 2002).
- Знание «первичной» последовательности является началом процесса узнавания информации, которую кодирует эта последовательность.



**Каждый человек отличается от другого на 1 нуклеотид из 1000**

## Результаты анализа 1 хромосомы человека

- **Общее число локусов 137426**
- **Частота встречаемости – 1 ОНП на 1000 нуклеотидов**
- **1 ОНП 93% генов**
- **>5 ОНП >59% генов**
- **>10 ОНП 39% генов**
- **В среднем каждый ген содержит 4 ОНП**



# Методы



- ПЦР, ПДРФ (RFLP), ОНКП (SSCP, single-stranded conformation polymorphism), ТГГЭ (TGGE/DGGE, temperature or denaturing gradient gel electrophoresis), секвенирование
- ДНК-чипы. На стекло размером 3.6 см<sup>2</sup> наносится 30 000 образцов. Методы: обратная транскрипция мРНК экстрактов, детекция флюоресценции лазер-аргонным сканером или изотопными методами, определение свечения или радиоактивности под конфокальным микроскопом и компьютерный анализ данных.
- Чипы на основе олигонуклеотидов (GeneChip®, Affymetrix), в 1.6 см<sup>2</sup> - 65-400000 олигонуклеотидов.



**ОНП** - однонуклеотидный полиморфизм

**SNP** - single nucleotide polymorphism

Определяет **90%** вариабельности генома

**регуляторный**

**рОНП**

в промоторах,

в интронах,

в 3'-нетранслируемых

областях

**кодирующий**

**кОНП**

в кодирующей части генов

**β-Талассемия** - наследственное заболевание, при котором нарушается синтез одной или нескольких субъединиц (цепей) гемоглобина.

Нонсенс-мутация, определяется заменой **AAG** на **TAG** в 17 кодоне гена бета-глобина, обнаружена в 1979 г. Чангом и Каном



# Классификация мутаций по типу изменения нуклеотидной последовательности



## Транзиции

70%

Пурин Пурин  
С→Т    G→A

## Трансверсии

30%

Пурин    Пиримидин  
С→A    G→Т  
С→G    G→С  
Т→A    A→Т



# Классификация мутаций по функциональному действию



**Мутации  
сдвига рамки  
считывания**

**Механизм:  
делеции или инсерции,  
некратные 3 нуклеотидам**

**Действие:  
образование  
нефункционального белка**

**Образование укороченного  
белка**

**Миссенс-мутации**

**70% не влияют на функцию**

**Классифицируются по  
положению:  
- 1 или 2 положения**

**Вырожденность генетического  
кода**

**CGA -кодируют  
CGG аргинин  
CGC  
CGU**

**Этот тип мутаций  
в 30% случаев  
приводит к активации  
сайтов сплайсинга**

**Нонсенс-мутации**

**Замена  
основания  
приводит к  
образованию  
одного из 3  
стоп-кодонов**

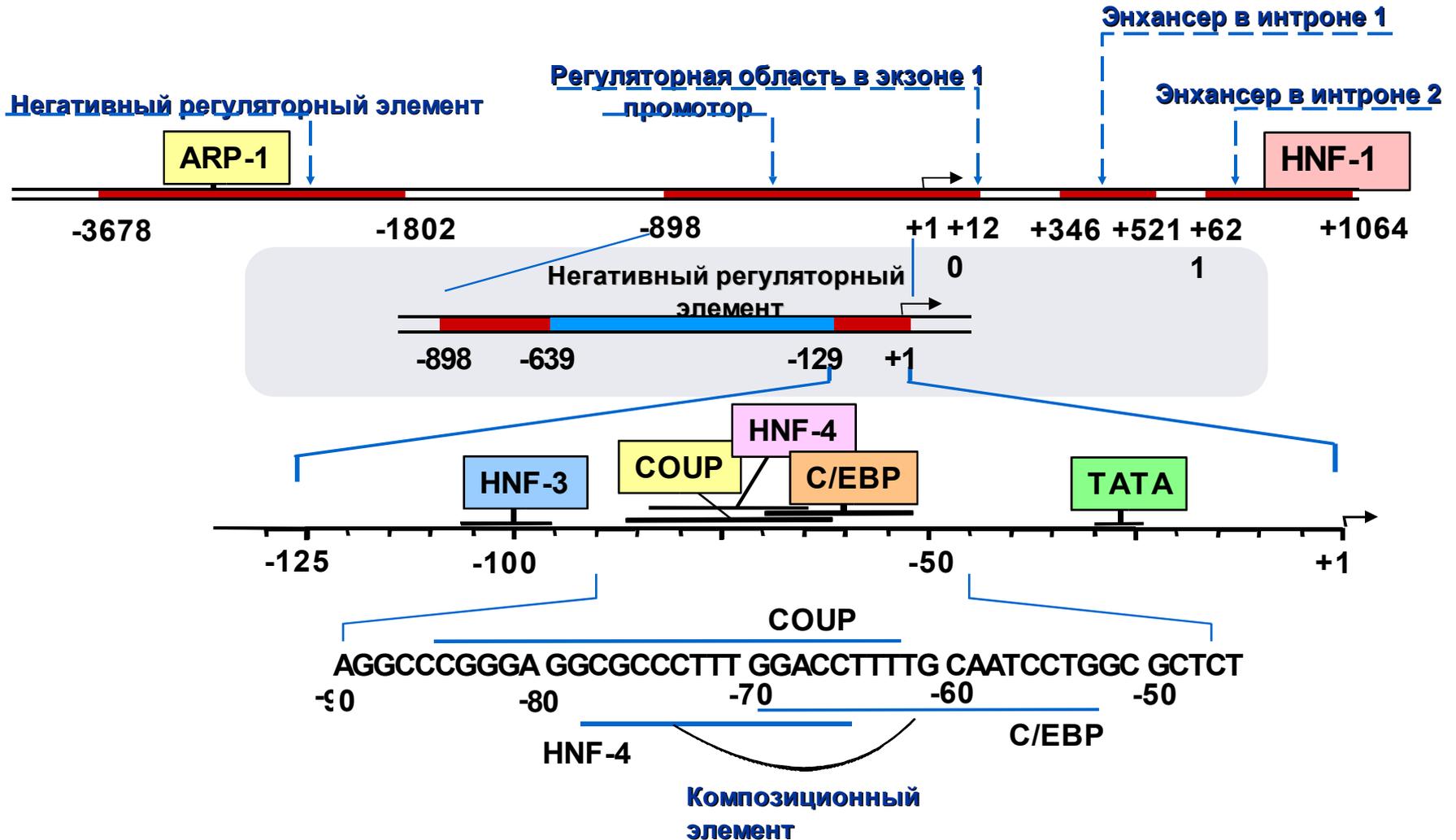
**UAA  
UAG**

**UGA**

**Терминация синтеза  
белка**

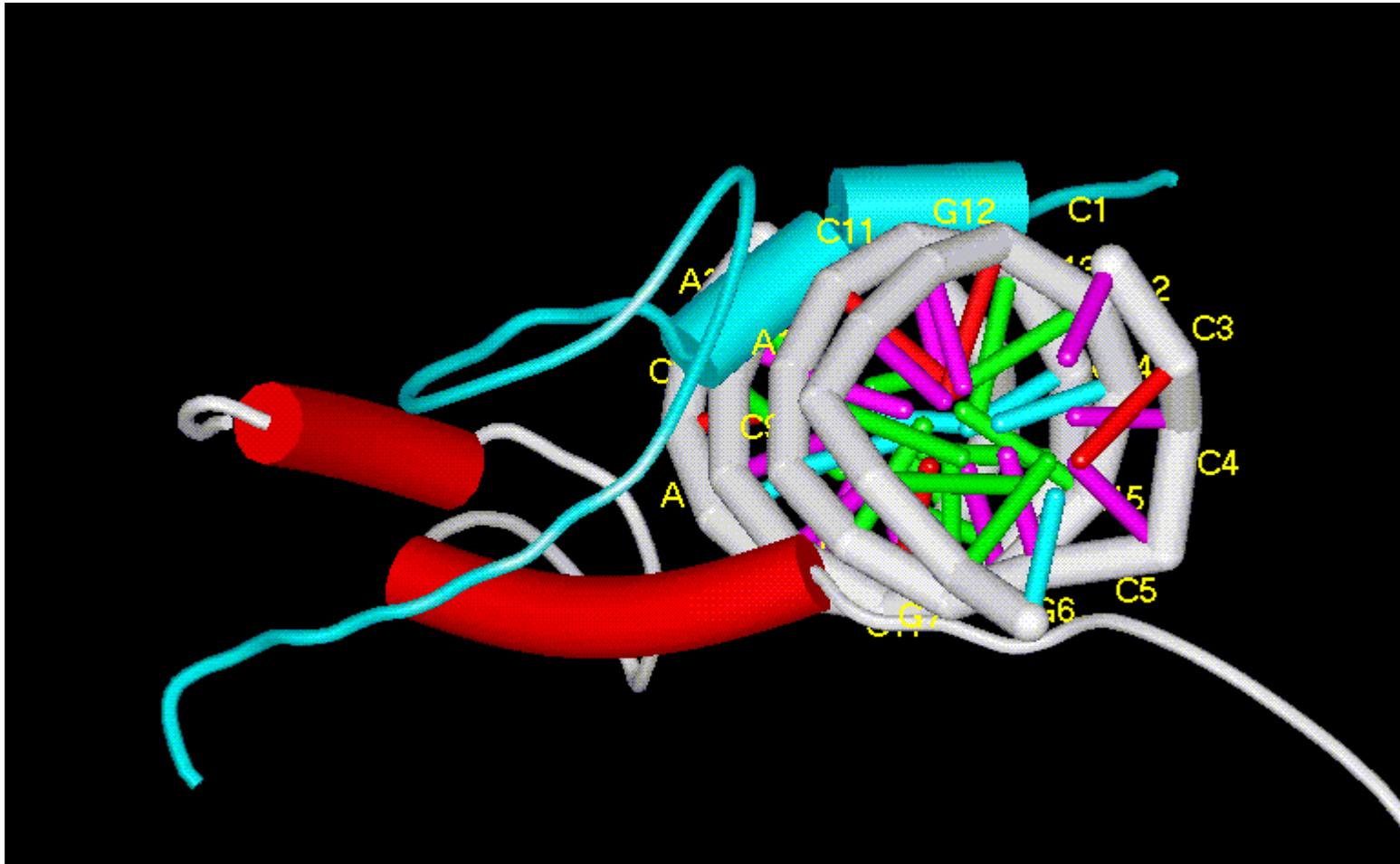


# База данных TRRD: организация регуляторных районов гена человека, кодирующего белок аполипопротеин В





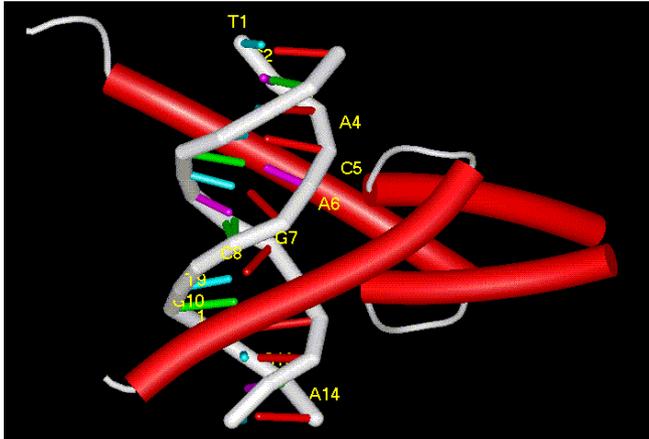
# Моделирование влияния мутаций: трехмерная структура ДНК-белкового комплекса



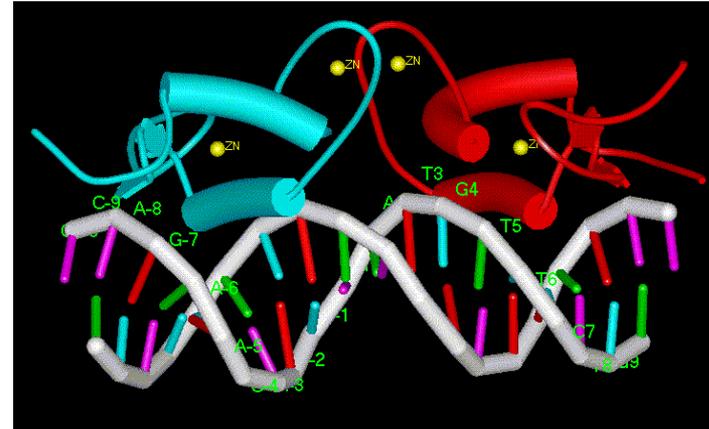
[http://www.mgs.bionet.nsc.ru/mgs/programs/acts2/images/USF\\_X.html](http://www.mgs.bionet.nsc.ru/mgs/programs/acts2/images/USF_X.html)



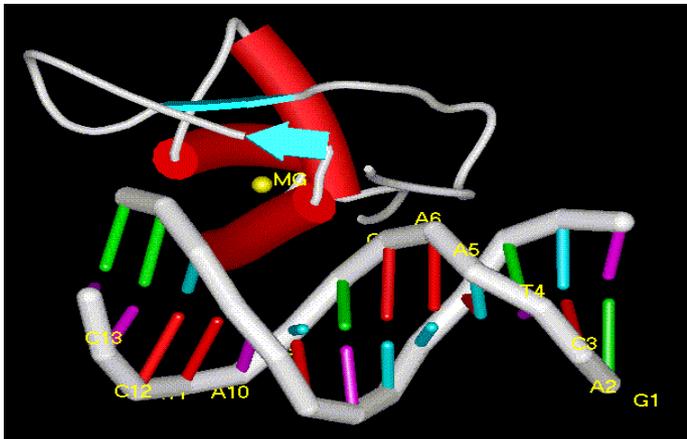
## I MyoD-DNA complex



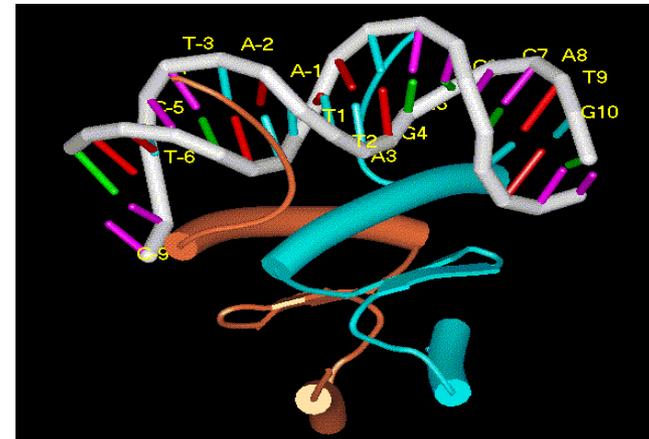
## II GR-DNA complex



## III HNF3-DNA complex



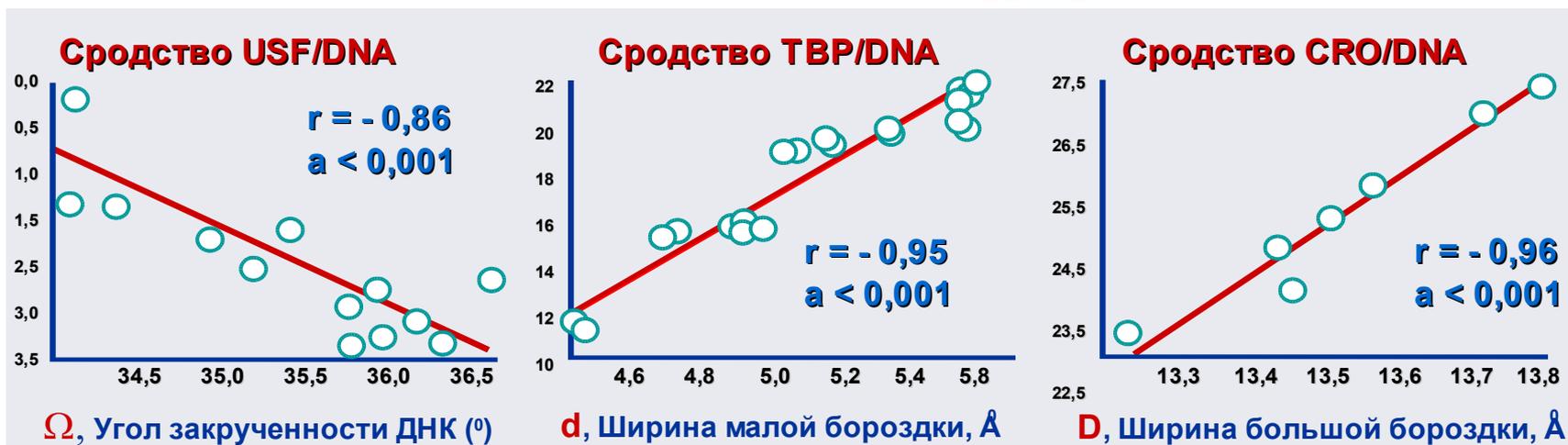
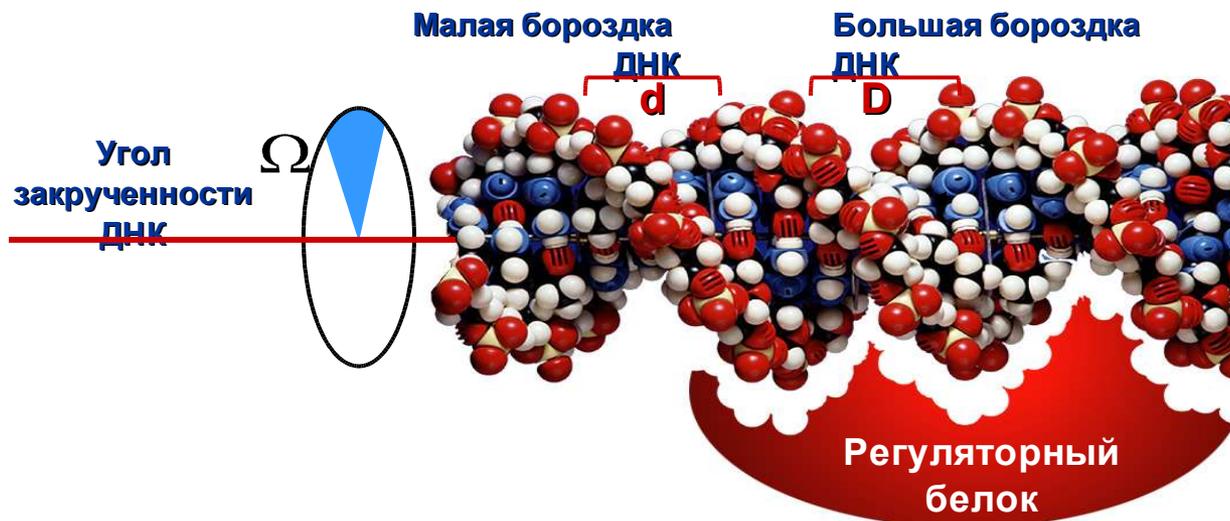
## IV SRF-DNA complex





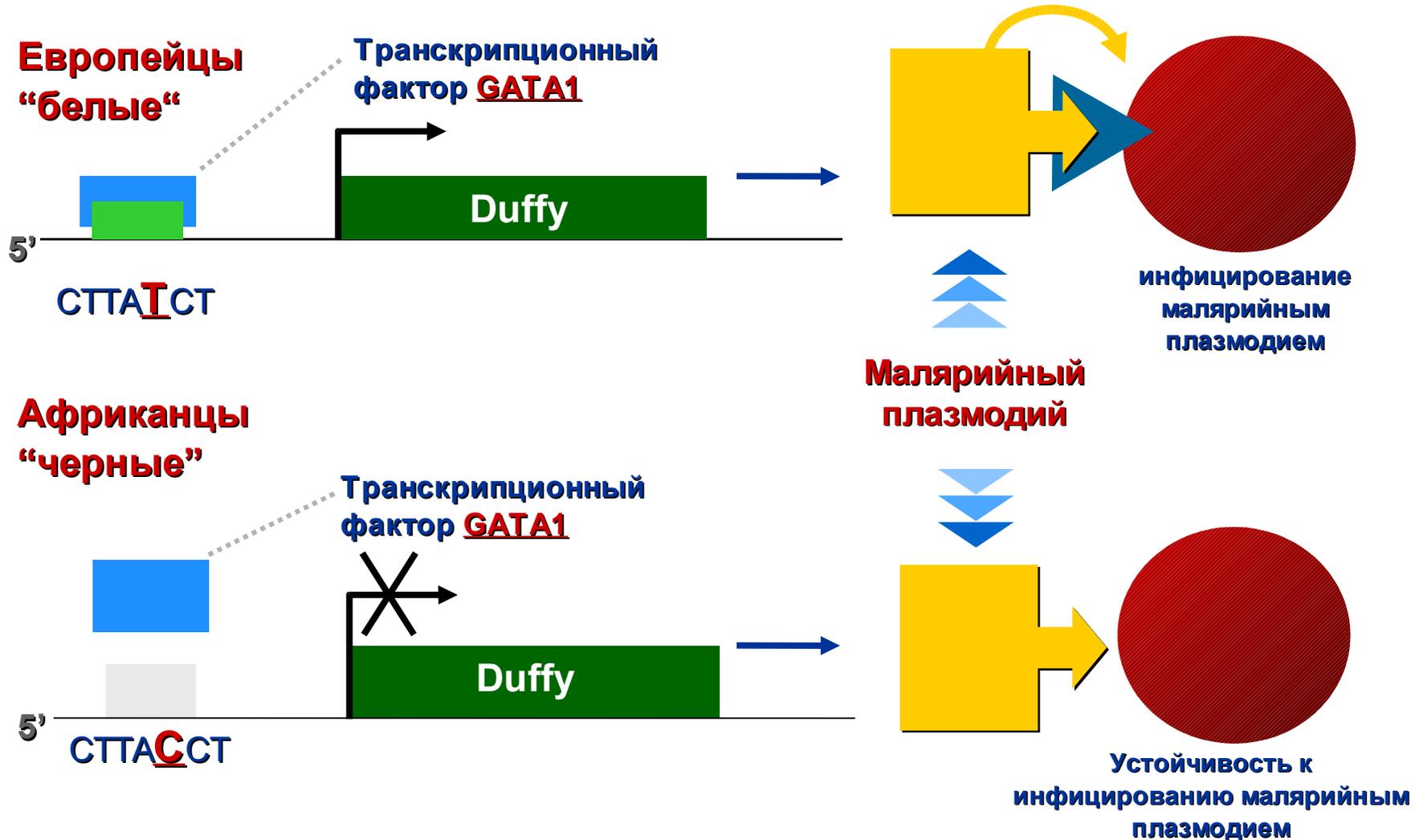


# Моделирование влияния мутаций: средство регуляторных белков к сайтам их связывания определяется конформационными свойствами ДНК





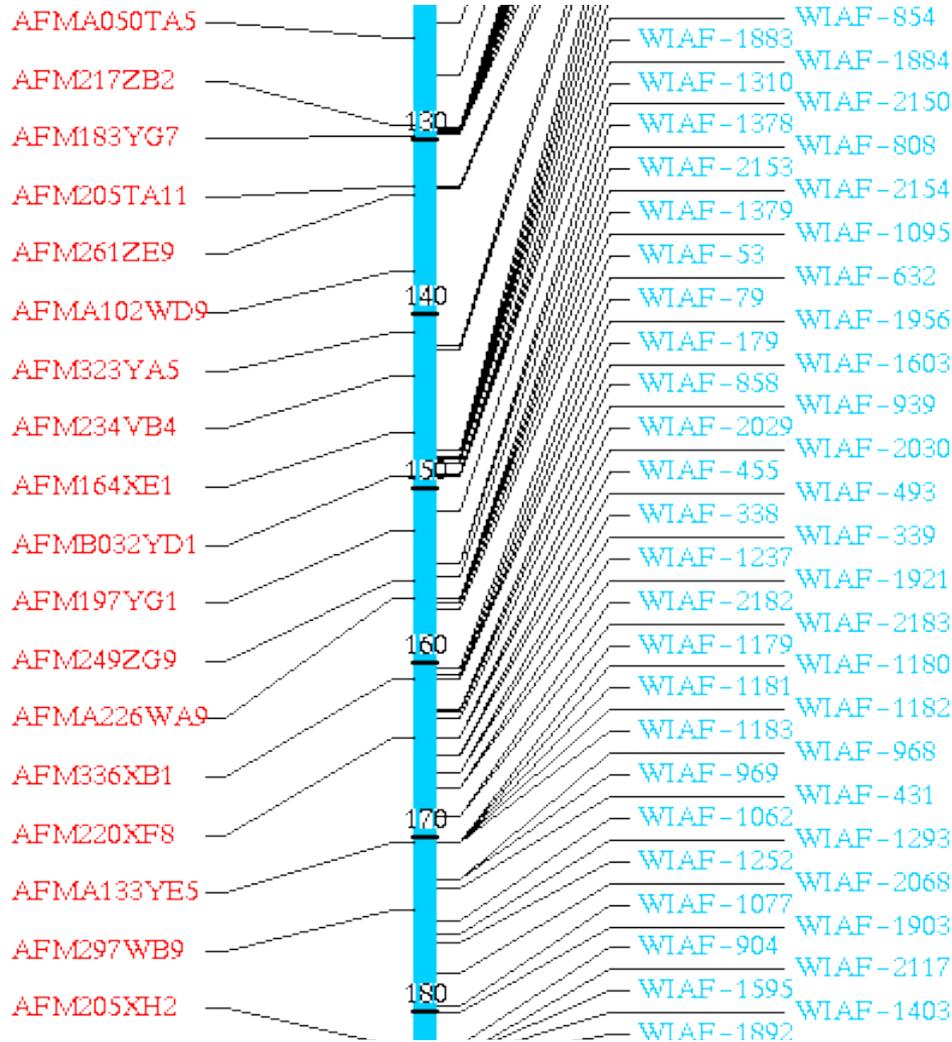
# Влияние мутаций на регуляторные районы генов полиморфизм по регуляторному району гена **DUFFY**





# Human SNP Database

<http://www-genome.wi.mit.edu/snp/human/>



21 октября 2002 – 13 релиз  
базы данных содержит  
193,037 ОНП; для 78,570 из  
которых приводятся  
аллельные частоты



## Обозначения мутаций в базах данных. Уровень ДНК



- Замещения обозначаются знаком >
- Запись 1997G>T означает, что в позиции 1997 нуклеотид G заменен на T.
- Делеции обозначаются “del”, запись 1997-1999-delTTC обозначает, что нуклеотиды TTC вырезаны из последовательности, начиная от 1997 до 1999 нуклеотида
- Инсерции обозначаются “ins”. Запись 1997-1998insT обозначает, что нуклеотид T вставлен между нуклеотидами 1997 и 1998
- Полиморфизм коротких повторов: AСТGTGTGCC (A- 1991 nt) обозначается как 1993(TG)3-22, то есть, в позиции 1993 содержится первый TG нуклеотид, который может быть повторен от 3 до 22 раз
- Мутации в интронах: запись IVS4-2A>C(1998-2A>C) обозначает, что замена A на C в положении -2 от начала интрона 4, или на уровне кДНК между нуклеотидами 1997 и 1998. IVS1+1G>C обозначает мутацию в донорном сайте сплайсинга, а IVS1-2A>G, в акцепторном сайте сплайсинга.



## Обозначения мутаций в базах данных. Уровень белка и РНК

- T97-C102del или T97-C102del6 означает, что аминокислоты от Треонина в положении 97 до Цистеина в положении 102 делетированы.
- Стоп кодон обозначается как X. Запись R97X означает замену Аргинина 96 на кодон терминации транскрипции.
- Запись R97fs означает мутацию сдвига рамки считывания, начинающуюся с Аргинина 97. Иногда обозначают длину сдвинутой рамки: R97fsX191, означает, что при мутации, новая рамка считывания открыта для 23 аминокислот.
- Запись r.15c>u означает, что она относится к РНК последовательности
- Варианты сплайсинга: [16g>t +16\_112del], для РНК обозначаются прописными буквами, запись обозначает мутацию, влияющую на сплайсинг и приводящую к образованию двух вариантов мРНК, одного нормального, а другого - с делецией нуклеотидов от 16 до 112.



Address <http://www.ncbi.nlm.nih.gov/SNP/> Go Links >>

**Single Nucleotide Polymorphism**

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search  for  Go Clear

Limits Preview/Index/History/Clipboard/Details

**dbSNP BUILD 108**

**GENERAL**  
dbSNP Home Page  
SNP Science Primer  
Announcements  
dbSNP Summary  
FTP SERVER  
Getting Started **NEW**  
Build History  
Handle Request

**DOCUMENTATION**  
FAQ  
Overview  
How To Submit  
RefSNP Summary Info  
Database Schema  
html  
pdf  
Data formats  
Heterozygosity computation

**SEARCH**  
Entrez SNP **NEW**  
Blast SNP  
Batch query  
By Submitter  
New Batches

**dbSNP Search Options**

Entrez SNP	ID Numbers	Submission Info	Batch	Locus Info	Free Form	Easy Form	Between Markers
------------	------------	-----------------	-------	------------	-----------	-----------	-----------------

**ANNOUNCEMENT**

NCBI has moved all FTP services to a new address: <ftp.ncbi.nih.gov>. The full contents of dbSNP can now be found at the new address <ftp://ftp.ncbi.nih.gov/snp/>. Please contact [snp-admin@ncbi.nlm.nih.gov](mailto:snp-admin@ncbi.nlm.nih.gov) to report problems with access to the new ftp area.

**Search by IDs**

Note: rs# and ss# must be prefixed with "rs" or "ss", respectively (ie. rs25, ss25)

Reference cluster ID(rs#)

Search Reset

[Advance ID Search](#)

**Submission Information**

[By Submitter](#)  
[New Batches](#)  
[Method](#)  
Population  
- [Detail](#) (Description, Handle, and ID)  
- [Class](#) (Classification based on geographic



# HGVbase

Выпуск 13.0:

1,451,426 входов (27

марта 2002)

The screenshot shows the EMBL-EBI SNP-Fasta Submission Form. The browser address bar shows <http://www.ebi.ac.uk/snpfasta3/>. The page header includes the EMBL-EBI logo and navigation links: EBI Home, About EBI, Research, Services, Toolbox, Databases, Downloads, Submissions. The main heading is "SNP-Fasta Submission Form".

Below the heading is a descriptive paragraph: "SNPs (Single Nucleotide Polymorphisms) comprise the largest set of sequence variants in most organisms. SNPs are useful polymorphic markers for mapping and discovering the genes associated with common diseases, or genes related to drug responsiveness. Search SNP sequences with the Fasta server."

The form contains several sections:

- YOUR EMAIL:** An empty text input field.
- SEARCH TITLE:** A text input field containing "Sequence".
- RESULTS:** A dropdown menu set to "interactive".
- PROGRAM:** A dropdown menu set to "fasta3".
- DATABASES:** Two buttons: "Nucleic Acid" and "HGVbase".
- GAP PENALTIES:** A dropdown menu set to "-12".
- SCORES & ALIGNMENTS:** Two dropdown menus: "SCORES" set to "50" and "ALIGN" set to "50".
- KTUP/HISTOGRAM:** Two dropdown menus: "KTUP" set to "6" and "HIST" set to "no".
- DNA STRAND:** A dropdown menu set to "none".
- MATRIX:** A dropdown menu set to "none".
- EXPECTATION UPPER VALUE:** A dropdown menu set to "1.0".
- EXPECTATION LOWER VALUE:** A dropdown menu set to "default".
- SEQUENCE RANGE:** A dropdown menu set to "START-END".
- DATABASE RANGE:** A dropdown menu set to "START-END".
- MOLECULE TYPE:** A dropdown menu set to "DNA".

Below these sections is a large text area for "Enter or Paste a DNA/RNA Sequence in any format:" with a "Help" button. At the bottom, there is an "Upload a file:" section with a "Browse..." button, and "Run" and "Reset" buttons.





**ALFRED**  
Allele Frequency Database

Data last modified 10/8/2002

Ongoing funding of ALFRED is provided by NSF grant BCS0096588.  
Initial funding for ALFRED was provided by NSF grant SBR-9632509 and USPHS grants P01GM57672, R01AA09379, and T15LM07056.

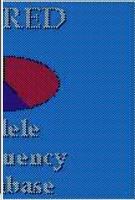
Information	Search	Summaries
<a href="#">Example Searches</a>	<a href="#">Basic Search</a>	<a href="#">Sites List</a>
<a href="#">About ALFRED</a>	<a href="#">Loci</a>	<a href="#">Table Numbers</a>
<a href="#">Contact</a>	<a href="#">Populations</a>	

 Do you have a favorite polymorphism you think should be included in ALFRED? [Email us your suggestions](#).

We are currently attempting to assemble the widely diffuse (in the literature) frequency data on ACE, CD4, PLAT and DM1.  
If you have published (or unpublished) data or know of a publication not yet included in ALFRED, please help us by [contacting us](#).

**News:**  
9/15/02 - A new graphical overview of the contents in ALFRED can be seen under [Table Numbers](#).  
6/25/02 - ALFRED was down this morning for a hardware upgrade.  
4/25/02 - We have begun entering data from the literature. Please advise us of any problems you may encounter.  
1/8/02 - Publication information is now output for populations.  
12/5/01 - The "Typing Summary" page which listed the number of individuals typed for a sample and site was removed as it had grown too large.  
10/28/01 - Some links to other databases have been corrupted due to changes in their address structure. We are working on fixing these

Internet



**ALFRED**  
Allele Frequency Database

Data last modified 10/8/2002

Ongoing funding of ALFRED is provided by NSF grant BCS0096588.  
Initial funding for ALFRED was provided by NSF grant SBR-9632509 and USPHS grants P01GM57672, R01AA09379, and T15LM07056.

Information	Search	Summaries
<a href="#">Example Searches</a>	<a href="#">Basic Search</a>	<a href="#">Sites List</a>
<a href="#">About ALFRED</a>	<a href="#">Loci</a>	<a href="#">Table Numbers</a>
<a href="#">Contact</a>	<a href="#">Populations</a>	

 Do you have a favorite polymorphism you think should be included in ALFRED? [Email us your suggestions](#).

We are currently attempting to assemble the widely diffuse (in the literature) frequency data on ACE, CD4, PLAT and DM1.  
If you have published (or unpublished) data or know of a publication not yet included in ALFRED, please help us by [contacting us](#).

**News:**  
9/15/02 - A new graphical overview of the contents in ALFRED can be seen under [Table Numbers](#).  
6/25/02 - ALFRED was down this morning for a hardware upgrade.  
4/25/02 - We have begun entering data from the literature. Please advise us of any problems you may encounter.  
1/8/02 - Publication information is now output for populations.  
12/5/01 - The "Typing Summary" page which listed the number of individuals typed for a sample and site was removed as it had grown too large.  
10/28/01 - Some links to other databases have been corrupted due to changes in their address structure. We are working on fixing these

Internet



## National Center for Biotechnology Information

### OMIM Locus-specific mutation database links

OMIM has links to a number of locus specific mutation databases. Other listings of such databases may be found at [HGMD](#) and the [HUGO Mutation Database Initiative](#).

- [The Androgen Receptor Mutations Database](#)
- [The Ataxia-Telangiectasia Database](#)
- [The BIOMDB database of mutations causing tetrahydrobiopterin deficiencies](#)
- [The ETKBase mutation registry for X-Linked agammaglobulinemia](#)
- [The Cystic Fibrosis Mutation Database](#)
- [The Emery-Dreifuss Muscular Dystrophy Mutation Database](#)
- [The Emory University MitoMap mitochondrion genome database](#)
- [The Factor VII Mutation Database](#)
- [The Fanconi Anemia Mutation Database](#)
- [The Favism Database of Glucose-6-Phosphate Mutations](#)
- [The Glycogen Storage Disease Type II \(Pompe Disease\) Mutation Database](#)
- [The Hemochilia A Mutation Database](#)
- [The Hereditary Non-Polyposis Colorectal Cancer Database](#)
- [The Hexosaminidase A Locus Database](#)
- [The Human Type I and Type III Collagen Mutation Database](#)
- [The IL2RG mutation database for X-linked SCID](#)
- [The Iowa Compendium of Rhodopsin and RDS Mutations](#)
- [The L1CAM Mutation Database](#)
- [The LDL Receptor Mutation Database](#)
- [The Ornithine Transcarbamylase Structure and Mutation Database](#)
- [The p53 mutation database](#)
- [The PAX6 Mutation Database](#)

Internet

**Welcome to the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff.**

Copyright © University of Wales College of Medicine 1997-2002. All Rights Reserved.

**HGMD Search   Statistics   What's new   HGMD Background**

- [HGMD publications](#)
- [Newly added genes](#)
- [Locus-specific databases](#)
- [Electronic mutation submission](#)
- [Comment form](#)
- [Supplementary information](#)
- [Other useful links](#)
- [Mutation nomenclature](#)
  - [publications](#)

This database is maintained by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips and M.E. Mort with the assistance of N.S.T. Thomas & S. Abeysinghe. We gratefully acknowledge the support of:

**Pfizer   MACMILLAN   SUN LIFE**

**<<STOP PRESS>>**

HGMD enters into collaboration with Celera Genomics. Please see our [What's New](#) page for the full text of the announcement.

If you refer to HGMD in any publication, please cite Krawczak M, Cooper DN (1997) The Human Gene Mutation Database. *Trends Genet.* 13:121-122. Articles describing the HGMD facility have also appeared in *Nucleic Acids Res.* 26: 285-287, 1998 and *Human Mutation* 15: 45-51, 2000.

For a picture of how others view HGMD, [click here](#).

Internet

Data last modified 10/8/2002

Ongoing funding of ALFRED is provided by NSF grant ECS0096588. Initial funding for ALFRED was provided by NSF grant SBR-9632509 and USPHS grants P01GM57672, R01AA09579, and T15LM07056.

**Information   Search   Summaries**

**Example Searches   Basic Search   Sites List**

**About ALFRED   Loci   Table Numbers**

**Contact   Populations**

Do you have a favorite polymorphism you think should be included in ALFRED? Email us your suggestions. We are currently attempting to assemble the widely diffuse (in the literature) frequency data on ACE, CD4, PLAT and DM1. If you have published (or unpublished) data or know of a publication not yet included in ALFRED, please help us by [contacting us](#).

**News:**

- 9/15/02 - A new graphical overview of the contents in ALFRED can be seen under [Table Numbers](#).
- 6/25/02 - ALFRED was down this morning for a hardware upgrade.
- 4/25/02 - We have begun entering data from the literature. Please advise us of any problems you may encounter.
- 1/8/02 - Publication information is now output for populations.
- 12/5/01 - The "Typing Summary" page which listed the number of individuals typed for a sample and site was removed as it had grown too large.
- 10/28/01 - Some links to other databases have been corrupted due to changes in their address structure. We are working on fixing these.

Internet



## Universal Mutation Database

Software and databases for mutations in human genes

### UNIVERSAL MUTATION DATABASE SOFTWARE

UMD softwares to query on line the APC, FBN1, LDLR, p53, VHL and WT1 databases

### DISEASE OR GENE SPECIFIC DATABASES

Links to gene/disease specific web sites (Marfan, LDLR, Myotonic Dystrophy and p53 web sites)

### OTHER LOCUS SPECIFIC DATABASES

Links to other locus specific databases around the world

### MULTI-GENES DATABASES

Links to databases including more than one gene

### CORE DATABASES

Links to the major databases (EBI, GDB, GenAtlas, HGMD, HUGO and NCBI)

Welcome to the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff.

Copyright © University of Wales College of Medicine 1997-2002. All Rights Reserved.

**HGMD**  
The Human Gene Mutation Database Project

**CELERA**

[HGMD Search](#) [Statistics](#) [What's new](#) [HGMD Background](#)

- [HGMD publications](#)
- [Newly added genes](#)
- [Locus-specific databases](#)
- [Electronic mutation submission](#)
- [Comment form](#)
- [Supplementary information](#)
- [Other useful links](#)
- [Mutation nomenclature
  - \[publications\]\(#\)](#)

This database is maintained by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips and M.E. Mort with the assistance of N.S.T. Thomas & S. Abeysinghe. We gratefully acknowledge the support of:

**Pfizer** **MACMILLAN** **SUN LIFE**

**<<STOP PRESS>>**

HGMD enters into collaboration with Celera Genomics. Please see our What's New page for the full text of the announcement.

If you refer to HGMD in any publication, please cite Krawczak M, Cooper DN (1997) The Human Gene Mutation Database. *Trends Genet.* 13:121-122. Articles describing the HGMD facility have also appeared in *Nucleic Acids Res.* 26: 285-287, 1998 and *Human Mutation* 15: 45-51, 2000.

For a picture of how others view HGMD, click here.

**Allele Frequency Database**

Data last modified 10/8/2002

Ongoing funding of ALFRED is provided by NSF grant ECS0096588  
Initial funding for ALFRED was provided by NSF grant SBR-9632509 and USPHS grants R01GM57672, R01AA09579, and T15LM07056

<a href="#">Information</a>	<a href="#">Search</a>	<a href="#">Summaries</a>
<a href="#">Example Searches</a>	<a href="#">Basic Search</a>	<a href="#">Sites List</a>
<a href="#">About ALFRED</a>	<a href="#">Loci</a>	<a href="#">Table Numbers</a>
<a href="#">Contact</a>	<a href="#">Populations</a>	

Do you have a favorite polymorphism you think should be included in ALFRED? Email us your suggestions. We are currently attempting to assemble the widely diffuse (in the literature) frequency data on ACE, CD4, PLAT and DM1. If you have published (or unpublished) data or know of a publication not yet included in ALFRED, please help us by contacting us.

**News:**

- 9/15/02 - A new graphical overview of the contents in ALFRED can be seen under Table Numbers.
- 6/25/02 - ALFRED was down this morning for a hardware upgrade.
- 4/25/02 - We have begun entering data from the literature. Please advise us of any problems you may encounter.
- 1/8/02 - Publication information is now output for populations.
- 12/5/01 - The "Typing Summary" page which listed the number of individuals typed for a sample and site was removed as it had grown too large.
- 10/28/01 - Some links to other databases have been corrupted due to changes in their address structure. We are working on fixing these.



  
**The Androgen Receptor Gene Mutations Database World Wide Web Server**

**The Lady Davis Institute for Medical Research,  
 Sir Mortimer B. Davis Jewish General Hospital**



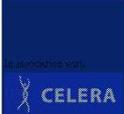
Last updated: 5/09/2002

Welcome to the Androgen Receptor Gene Mutations Database World Wide Web Server. This server is for all who are interested in mutations of the Androgen Receptor Gene. Please feel free to send your feedback to [bruce.gottlieb@mcgill.ca](mailto:bruce.gottlieb@mcgill.ca)

- [Introduction to androgen receptor gene mutations](#)
- [NEW pdf version of Database of Androgen Receptor Gene Mutations.](#)  
 NB: This database is updated every two weeks
- Phenotypes in **orange** indicate a somatic mutation.
- [Table of Androgen Receptor-Interacting Proteins \(new version 15/3/02\)](#)
- [Map of Androgen Receptor-Interacting Proteins \(new version 20/12/00\)](#)
- [References for Androgen Receptor-Interacting Proteins \(new version 15/03/02\)](#)
- [Regions of interaction of Androgen Receptor-Interacting Proteins \(new version 15/03/02\)](#)
- Additional information from the database is available as [Selectable Options](#)

 **Welcome to the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff.**

Copyright © University of Wales College of Medicine 1997-2000. All Rights Reserved.



[HGMD Search](#)   [Statistics](#)   [What's new](#)   [HGMD Background](#)

- [HGMD publications](#)
- [Newly added genes](#)
- [Locus-specific databases](#)
- [Electronic mutation submission](#)
- [Comment form](#)
- [Supplementary information](#)
- [Other useful links](#)
- [Mutation nomenclature](#)
  - [publications](#)

This database is maintained by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips and M.E. Mort with the assistance of N.S.T. Thomas & S. Abeyasinghe. We gratefully acknowledge the support of:

**<<STOP PRESS>>**

HGMD enters into collaboration with [Celera Genomics](#). Please see our [What's New](#) page for the full text of the announcement

If you refer to HGMD in any publication, please cite Krawczak M, Cooper DN (1997) The Human Gene Mutation Database. *Trends Genet.* 13:121-122. Articles describing the HGMD facility have also appeared in *Nucleic Acids Res.* 26: 285-287, 1998 and *Human Mutation* 15: 45-51, 2000.

[For a picture of how others view HGMD, click here.](#)

Data last modified 10/8/2002

  
 Ongoing funding of ALFRED is provided by NSF grant BCS0096588.  
 Initial funding for ALFRED was provided by NSF grant SBR-9632509 and USPHS grants P01GM57672, R01AA09379, and T15LM07056.

<a href="#">Information</a>	<a href="#">Search</a>	<a href="#">Summaries</a>
<a href="#">Example Searches</a>	<a href="#">Basic Search</a>	<a href="#">Sites List</a>
<a href="#">About ALFRED</a>	<a href="#">Loc</a>	<a href="#">Table Numbers</a>
<a href="#">Contact</a>	<a href="#">Populations</a>	

Do you have a favorite polymorphism you think should be included in ALFRED? Email us your suggestions. We are currently attempting to assemble the widely diffuse (in the literature) frequency data on ACE, CD4, PLAT and DM1. If you have published (or unpublished) data or know of a publication not yet included in ALFRED, please help us by [contacting us](#).

**News:**

- 9/15/02 - A new graphical overview of the contents in ALFRED can be seen under [Table Numbers](#)
- 6/25/02 - ALFRED was down this morning for a hardware upgrade.
- 4/25/02 - We have begun entering data from the literature. Please advise us of any problems you may encounter.
- 1/8/02 - Publication information is now output for populations.
- 12/5/01 - The "Typing Summary" page which listed the number of individuals typed for a sample and site was removed as it had grown too large.
- 10/28/01 - Some links to other databases have been corrupted due to changes in their address structure. We are working on fixing these.





**Serologicals**  
Creating A Healthier World

## Human Serum Albumin Mutations

Residue	Amino acid chg		Codon change (lower case) (+ = known)	Geographical names and [references], in order of reports
	From	To		
-2	Arg	His	CgT -> CaT+	Lille [1], Polibauer, Somalia, Tokushima [2], Taipei [3], Fukuoka-2 [4], Varese [5], Wu Yang [6], Mayo, Komagane-2* [7]
-2@	Arg	Cys	cGT -> cGT+	[2 to 45 omitted] Malmo I [8], Fakoura [9], Trieste [5] Reddit (see also residue 320[10,11], high in Italy, Sweden, 3 homozygotes* [12], Lidat

**HGMD**  
The Human Gene Mutation Database

Welcome to the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff.

Copyright © University of Wales College of Medicine 1997-2000. All Rights Reserved.

In association with **CELERA**

[HGMD Search](#)   [Statistics](#)   [What's new](#)   [HGMD Background](#)

- [HGMD publications](#)
- [Newly added genes](#)
- [Locus-specific databases](#)
- [Electronic mutation submission](#)
- [Comment form](#)
- [Supplementary information](#)
- [Other useful links](#)
- [Mutation nomenclature](#)
  - [publications](#)

This database is maintained by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips and M.E. Mort with the assistance of N.S.T. Thomas & S. Aboysinghe. We gratefully acknowledge the support of:

**Pfizer**   **MACMILLAN**   **SUN LIFE**

<<STOP PRESS>>

HGMD enters into collaboration with **Celera Genomics**. Please see our [What's New](#) page for the full text of the announcement.

If you refer to HGMD in any publication, please cite Krawczak M, Cooper DN (1997) The Human Gene Mutation Database. *Trends Genet.* 13:121-122. Articles describing the HGMD facility have also appeared in *Nucleic Acids Res.* 26: 285-287, 1998 and *Human Mutation* 15: 45-51, 2000.

**McGill**

## The Androgen Receptor Gene Mutations Database World Wide Web Server

The Lady Davis Institute for Medical Research,  
Sir Mortimer B. Davis-Jewish General Hospital

Let updated: 1/09/2002

Welcome to the Androgen Receptor Gene Mutations Database World Wide Web Server. This server is for all who are interested in mutations of the Androgen Receptor Gene. Please feel free to send your feedback to [bruce.ponitub@mcgill.ca](mailto:bruce.ponitub@mcgill.ca)

- [Introduction to androgen receptor gene mutations](#)
- [NEW pdf version of Database of Androgen Receptor Gene Mutations.](#)  
NB: This database is updated every two weeks
- [Phenotypes in usage indicate a somatic mutation.](#)
- [Table of Androgen Receptor-Interacting Proteins \(new version 1/3/02\)](#)
- [Map of Androgen Receptor-Interacting Proteins \(new version 20/12/00\)](#)
- [References for Androgen Receptor-Interacting Proteins \(new version 1/3/02\)](#)
- [Regions of Interaction of Androgen Receptor-Interacting Proteins \(new version 1/3/02\)](#)
- [Additional information from the database is available as Selectable Options](#)

**HGMD**  
The Human Gene Mutation Database

Welcome to the Human Gene Mutation Database at the Institute of Medical Genetics in Cardiff.

Copyright © University of Wales College of Medicine 1997-2000. All Rights Reserved.

In association with **CELERA**

[HGMD Search](#)   [Statistics](#)   [What's new](#)   [HGMD Background](#)

- [HGMD publications](#)
- [Newly added genes](#)
- [Locus-specific databases](#)
- [Electronic mutation submission](#)
- [Comment form](#)
- [Supplementary information](#)
- [Other useful links](#)
- [Mutation nomenclature](#)
  - [publications](#)

This database is maintained by D.N. Cooper, E.V. Ball, P.D. Stenson, A.D. Phillips and M.E. Mort with the assistance of N.S.T. Thomas & S. Aboysinghe. We gratefully acknowledge the support of:

**Pfizer**   **MACMILLAN**   **SUN LIFE**

<<STOP PRESS>>

HGMD enters into collaboration with **Celera Genomics**. Please see our [What's New](#) page for the full text of the announcement.

If you refer to HGMD in any publication, please cite Krawczak M, Cooper DN (1997) The Human Gene Mutation Database. *Trends Genet.* 13:121-122. Articles describing the HGMD facility have also appeared in *Nucleic Acids Res.* 26: 285-287, 1998 and *Human Mutation* 15: 45-51, 2000.

[For a picture of how others view HGMD, click here](#)

**ALFRED**  
Allele Frequency Database

Data last modified 10/8/2002

Choosing funding of ALFRED is provided by NSF grant EGS0096588  
Initial funding for ALFRED was provided by NSF grant SBR-9622919 and USPHS grants E01GM57672, R01AA09379, and T15LM07056

[Information](#)   [Search](#)   [Summaries](#)

[Example Searches](#)   [Basic Search](#)   [Sites List](#)

[About ALFRED](#)   [LocI](#)   [Table Numbers](#)

[Contact](#)   [Populations](#)

Do you have a favorite polymorphism you think should be included in ALFRED? [Email us your suggestions](#)

We are currently attempting to assemble the widely diffuse (in the literature) frequency data on ACE, CDM, PLAT and DM1. If you have published (or unpublished) data or know of a publication not yet included in ALFRED, please help us by [contacting us](#).

**News:**

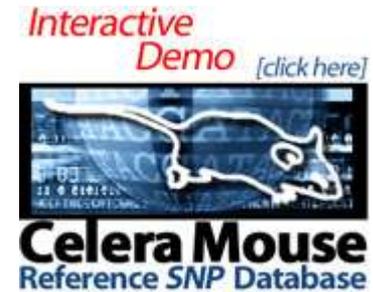
- 9/15/02 - A new graphical overview of the contents in ALFRED can be seen under [Table Numbers](#).
- 6/25/02 - ALFRED was down this morning for a hardware upgrade.
- 4/25/02 - We have begun entering data from the literature. Please advise us of any problems you may encounter.
- 1/8/02 - Publication information is now output for populations.
- 12/4/01 - The "Typing Summary" page which listed the number of individuals typed for a sample and site was removed as it had grown too large.
- 10/28/01 - Some links to other databases have been corrected due to changes in their address structure. We are working on fixing these.



# Использование модельных животных для изучения заболеваний человека



Инфекционные заболевания	Инбредные линии мыши	
	A/J	C57BL/6J
Пневмония легионеров	Восприимчивость +	Резистентность -
Малярия	+	-
Вирусный гепатит (MHV3)	-	+
Видоспецифический иммунодефицит (аналог СПИДа)	-	+
Мультифакториальные заболевания		
Артрит	+	-
Рак кишечника	+	-
Рак легкого	+	-
Астма	+	-
Атеросклероз	-	+



**Export variations for assay design**

**View location and effect of variation on gene structure**

**Link variation to identified gene of interest**

**Associate variation with strain information**

**View trace files for Celera sequenced variations to assess quality**



## Области применения ОНП



- **Изучение генетической предрасположенности к различным заболеваниям**
- **Изучение наследственной восприимчивости к действию различных лекарств**  
(скорость элиминации лекарства у разных пациентов может различаться от 4 до 40 раз)
- **Оценка риска возникновения мультифакторных заболеваний (около 300 ОНП позволяют оценить вероятность возникновения психоаффективных психозов, сахарного диабета, астмы, ожирения)**



# Гаплотипы



- ОНП наследуются сцепленно («блоками»)
- **Гаплотипы** - «предковые» хромосомные сегменты, которые не подверглись действию рекомбинации
- ОНП, входящие в гаплотип, тесно сцеплены и наследуются как отдельный блок
- Например, хромосома 21 содержит блок длиной 19000 нт, содержащий **26** ОНП; однако, 80% генетической изменчивости в популяции представлена всего **4** гаплотипами (Patil, 2001)
- Если бы эти ОНП наследовались независимо, наблюдали бы  **$2^{26}$**  аллельных комбинаций



## Обязательные компоненты:

- Возможность индивидуального генотип-специфического подбор фармацевтических препаратов для коррекции заболеваний человека.
- Выбор оптимальных стратегий коррекции индивидуальных генетических дефектов на основе компьютерного анализа и моделирования функции нарушенных генетически контролируемых систем и процессов организма.



# Один размер не подходит для всех: ФАРМАКОГЕНОМИКА



- В 1998 году, в Журнале Американской Медицинской Ассоциации опубликованы результаты исследования побочного действия лекарств в США в 1994 году:
  - > 2.2 миллиона серьезных заболеваний
  - > 100 000 смертельных случаев.
- В настоящее время, не существует универсального способа для надежного предсказания реакции пациента на лекарственное воздействие.
- Что такое Фармакогеномика?



# Терминология



- **Фармакогенетика** – наука, изучающая наследственные различия (полиморфизм) метаболизма лекарств и различия в ответе организма на действие лекарств.
- **Фармакогеномика** – наука, изучающая общие аспекты функционирования многих генов, детерминирующих реакцию на действие лекарств. **Фармакогеномика** изучает дифференциальную экспрессию генов, индуцированную под воздействием проверяемого вещества-агента в условиях *in vitro* или *in vivo*.
- На практике, эти два термина взаимозаменяемы.
- <http://www.ncbi.nlm.nih.gov/About/primer/pharm.html>



## **Задачи фармакогенетики**



- **Реакция пациента на лекарственную терапию вариабельна и часто непредсказуема – лекарство может действовать положительно, отрицательно или вообще не подействовать**
- **Лекарство, высоко эффективное для лечения одного пациента, может оказаться бесполезным для другого**
- **У некоторых пациентов могут развиваться побочные реакции**
- **Современный принцип фармакологии «один размер для всех» предназначен для лечения «среднего» пациента**



## Какова роль полиморфизма ДНК ?



- **Основная проблема:** какие варианты ДНК-последовательности ассоциированы с определенными наследуемыми фенотипами
- **Основной подход:** исследование взаимоотношений генотип-фенотип в различных популяциях
- **Задача фармакогенетики:** определить варианты генетической изменчивости (полиморфизмы), которые могут модифицировать ответ организма на действие лекарств



## Прочие проблемы:

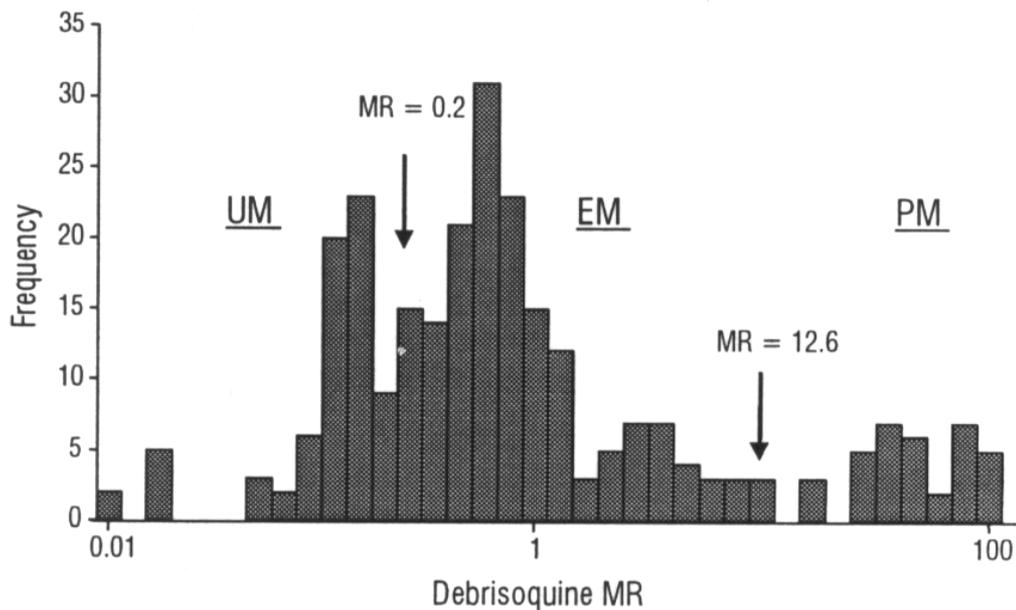


- **Повышение качества здравоохранения требует огромных вложений средств (разработка новых лекарств, генетическое тестирование)**
- **Этические проблемы подразделения пациентов на группы**
- **Проблемы частной собственности пациента на владение генетической информацией, конфиденциальность и права пользования этой информацией**



## Каковы источники первичных экспериментальных данных?

- Данные о генотипе
  - первичная последовательность ДНК
  - вариабельность последовательности ДНК у разных особей
- Фенотипические данные о клетках
  - дифференциальная экспрессия генов и протеомика
  - функциональный анализ
  - фармакокинетика и фармакодинамика
- Клинические данные



**Индекс метаболизма CYP2D6: распределение в популяции Швеции (Dahl ML et al., 1995)**

**MR скорость метаболизма дебрисоквина; UM - супер-активный метаболизм; EM - активный метаболизм (=норма); PM - слабый метаболизм.**

**MR=0.2**

**MR=12.6**



Мутации	Эффект	Норма		< функции		редукция функции										
		1	2A	2B	9	10A	10B	3	4A	4B	6A	6B	7	8	11	12
Систематическое название мутации <i>CYP2D6</i> *																
G119A	V11M															
C188T	P34S															
G212A	G42R															
G971C	Сплайсинг															
C1062A	L91M															
A1072G	H94R															
C1085G	Молчащая															
C1127T	Молчащая															
G1749C	Молчащая															
T1795del	Сдвиг рамки															
G1846T	Стоп кодон															
G1934A	Сплайсинг															
G2064A	G212E															
A2637del	Сдвиг рамки															
2701-3del	Делеция L281															
C2938T	R296C															
A3023C	H324P															
G4268C	S486T															
<b>Частота %</b>	Кавказская	34	28	5	2	3		1	20		1.5		<1	<1	<1	<1
	Только РМ	0	0	0	0	0		4	65		6		2	1	1	1

молчащая мутация

мутация, повреждающая 1 аминокислоту

мутация, приводящая к потере функциональности

(Meyer UA and Zanger UM, 1997 )



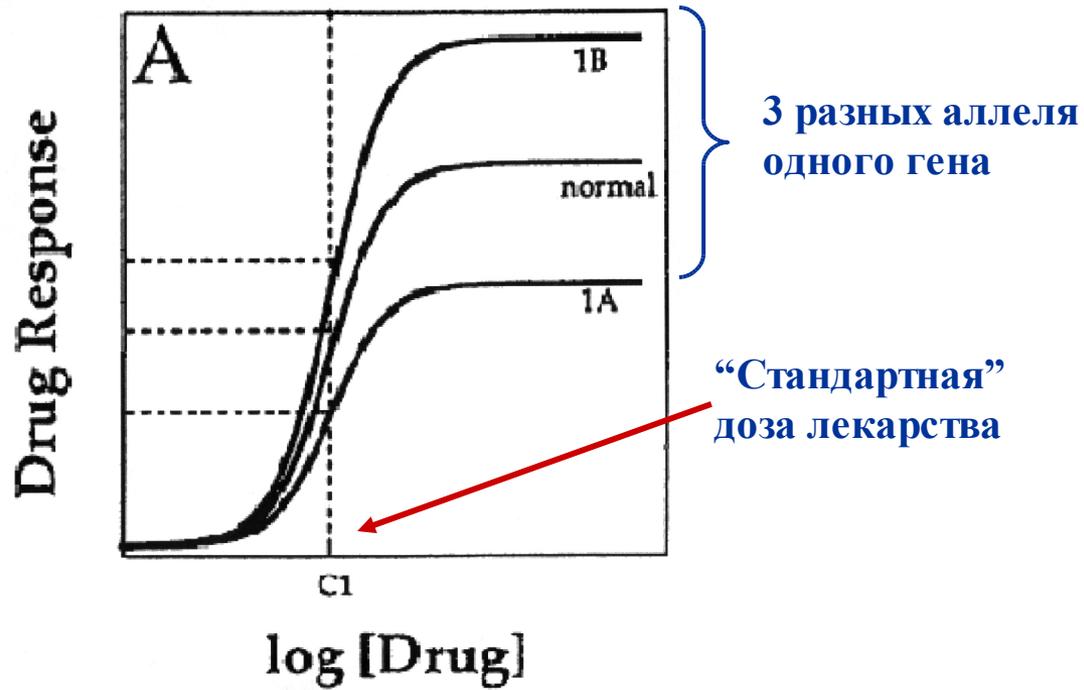
## Лекарства-аналоги пуринов



- **Класс лекарств (6-меркаптопурин, 6-тиогуанин, азатиопурин)**
- **Применяются для лечения лимфобластной лейкемии, аутоиммунных заболеваний, воспалений внутренних органов, и после трансплантации органов**
- **Действие: конкурируют с синтезом аминокислот**
- **Частично расщепляются ферментом тиопурин метилтрансферазой (ТРМТ)**
- **Терапевтическое действие ограничено супрессией клеток миелоидного ряда**



# Вариабельность ответа на введение лекарства

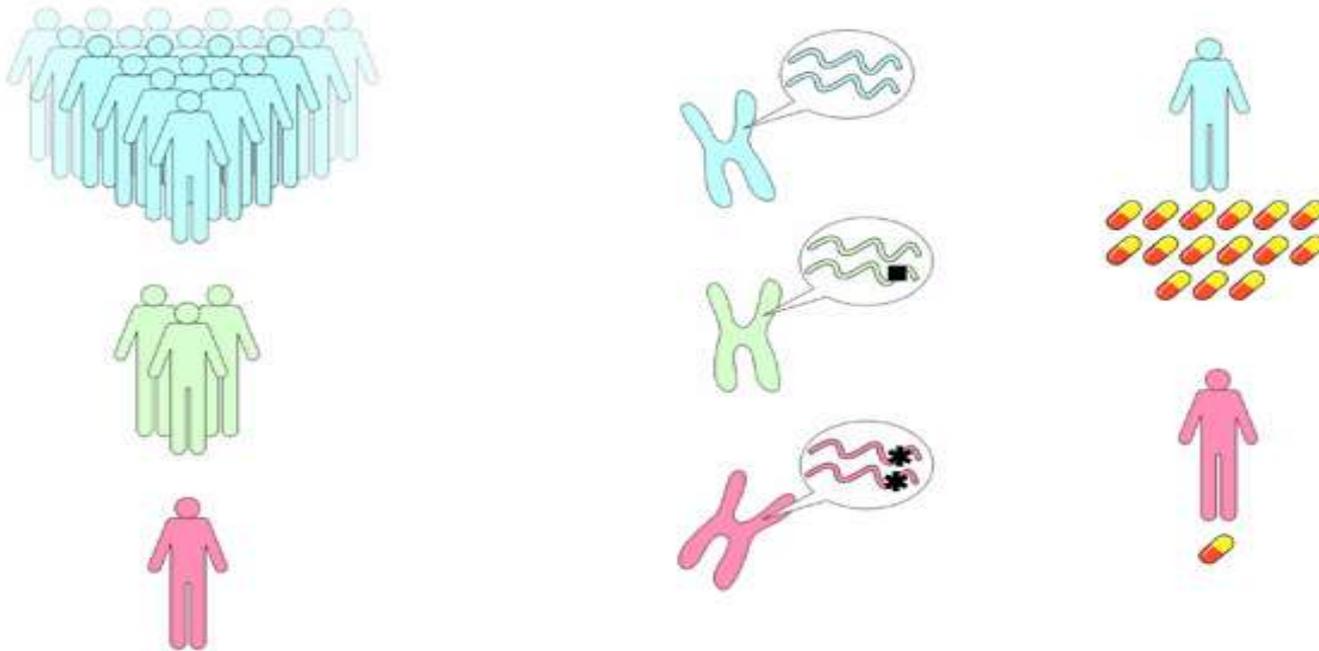




## Пациенты по-разному реагируют на введение противолейкемийного лекарства, 6-меркаптопурина.



Разный ответ обусловлен вариабельностью гена, кодирующего фермент ТРМТ, тиопурин метилтрансферазу.



Простой анализ крови поможет врачу назначить дозировку лекарства, соответствующую генетическому профилю больного (tailor-made medicine).

<http://www.nigms.nih.gov/news/reports/images/>

У большинства пациентов лекарство быстро выводится, поэтому для эффективной терапии лекарство следует назначать в больших дозах.

У части пациентов метаболизм этого лекарства замедлен, поэтому для лечения, а также для предотвращения токсического побочного действия, необходимы меньшие дозы.

У небольшого числа пациентов метаболизм настолько ослаблен, что действие лекарства может оказаться летальным.



## Клинический фенотип



- Построение **фармакокинетических профилей** (абсорбция, распространение и выведение лекарства)
- Построение **фармакодинамических профилей** (тестирование дыхательной функции, сердечной деятельности, физиологических показателей, оценка побочного действия)
- Оценка **клинического эффекта** (заболеваемость, смертность)



# Анализ молекулярного и клеточного фенотипа



- Молекулярный и клеточный функциональный анализ
- Анализ функциональных свойств белка по катализируемой им реакции (связывание, каталитическая активность, ингибирование, зависимость от концентрации лекарства)
- 3D структурные данные
- Скорость метаболизма лекарства в клетке
- Клеточная локализация
- Дифференциальная экспрессия генов



## Перспективы развития фармакогенетики



- Многие ОНП еще не документированы
- Функциональное действие многих ОНП пока не известно
- Неизвестна роль полиморфизма ДНК при проявлении мультифакторных заболеваний
- Лекарственный ответ определяется не только на генетическом уровне
- Необходимы новые подходы для хранения и обработки огромных массивов информации
- Часто неизвестна взаимосвязь между генотипом и фенотипом



# rSNP\_Guide, a database system for analysis of transcription factor binding to target sequences: application to SNPs and site-directed mutations

Julia V. Ponomarenko\*, Tatyana I. Merkulova, Gennady V. Vasiliev, Zoya B. Levashova, Galina V. Orlova, Sergey V. Lavryushev, Oleg N. Fokin, Mikhail P. Ponomarenko, Anatoly S. Frolov and Akinori Sarai<sup>1</sup>

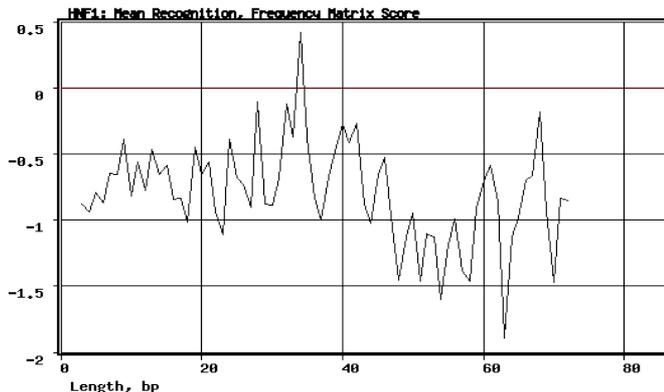
Institute of Cytology and Genetics, 10 Lavrentyev Avenue, Novosibirsk and Chemical Research, RIKEN, 3-1-1 Koyadai, Tsukuba, Japan

Received September 5, 2000; Revised and Accepted November 1, 2000

## ABSTRACT

rSNP\_Guide is a novel curated database system for analysis of transcription factor (TF) binding to target sequences in regulatory gene regions altered by mutations. It accumulates experimental data on naturally

representative high protein less silencing



rSNP-Tools - Microsoft Internet Explorer

Address: http://www.mgs.bionet.nsc.ru/mgs/programs/rsnp/

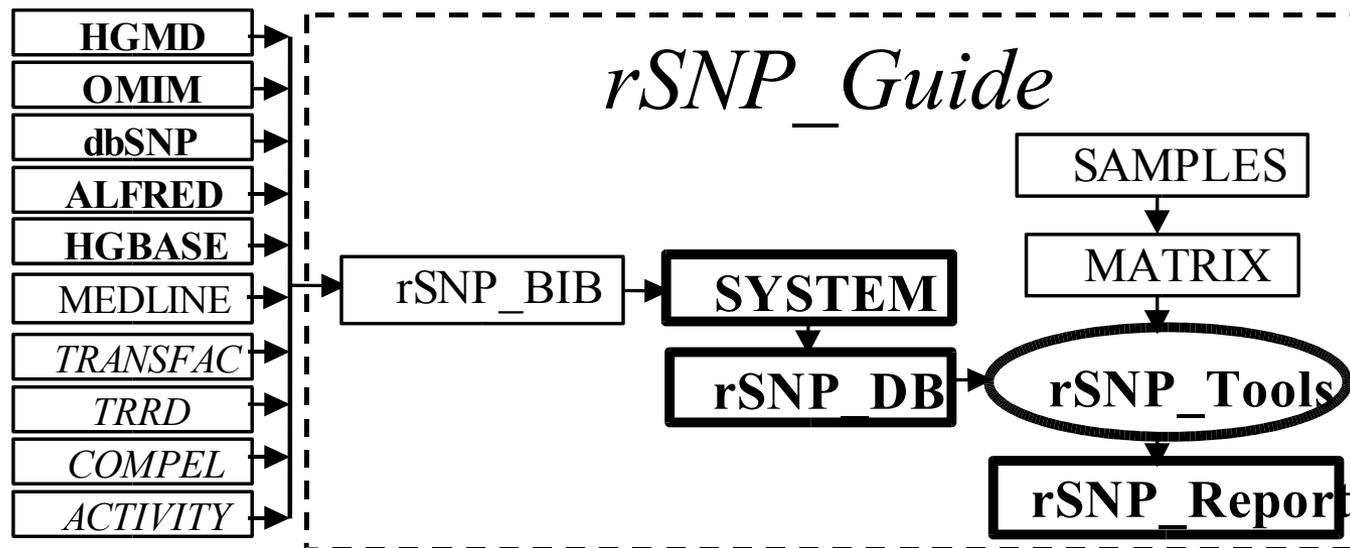
Select SNP's	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Statistical analysis				Significance Level <a href="#">[HELP]</a>			
Average of Distribution				0.93353	1.12794	0.91400	1.07311	0.025			
Mean Square Deviation				0.26713	0.13036	0.16163	0.29078	Accuracy <a href="#">[HELP]</a>			
								0.05			

Simulate the gel-shift data by assigning the X-site listed below [\[HELP\]](#)

DNA-protein Binding	SNP #1		SNP #2		SNP #3		Euclidean distances				Prediction <a href="#">[HELP]</a>
	(+)	(-)	(+)	(-)	(+)	(-)	X++	X+-	X-+	X--	
I X++	1	1	0	0			0	1.10453	1.10453	1.56204	MEF-2
II X+-	1	-0.1	0	-0.1			1.10453	0	1.56204	1.10453	
III X-+	-0.1	1	-0.1	0			1.10453	1.56204	0	1.10453	
X-site simulates the absence of DNA-protein complex											
IV X--	-0.1	-0.1	-0.1	-0.1			1.56204	1.10453	1.10453	0	TF-site is absent
No idea											
all (-)	OCT:										
Conflict											
Near zero	HNF1; HNF3:										

Calculate

Converter Report Clear All



Графическое представление основных компонентов системы *rSNP\_Guide*, предназначенной для анализа рОНП. Исходная информация экстрагируется из Интернет-доступных баз данных по мутациям: HGMD, dbSNP, HGBASE, ALFRED и OMIM; данные по искусственному мутагенезу взяты из баз данных TRANSFAC, TRRD, COMPEL и ACTIVITY. Используя оригинальные экспериментальные статьи (*rSNP\_BIB*), исходные экспериментальные условия документируются разделе *rSNP\_Guide*, **SYSTEM**. Данные по изменению паттернов связывания ядерных белков с последовательностью ДНК, содержащей ОНП, регистрируются в **rSNP\_DB**. Типичные примеры предсказания потенциального ТФ, связывающегося с мутантной последовательностью регистрируются при помощи Java-скрипт апплета *rSNP\_Tools*, который построен с использованием баз данных **SAMPLES** и **MATRIX**. Конечные результаты записываются в базу данных **rSNP\_Reports**.



## Примеры однонуклеотидного полиморфизма с известным клиническим эффектом в регуляторных районах генов



Заболевание	Ген	Район	ОНП	Последовательность ДНК	Ссылка
Восприимчивость к спонтанному и индуцированному мутагенезу в легких	<i>K-ras</i>	2 интрон 296 A/C 288 C/G 296 A/C	CA CC GC	5'-AAGAAA C TCCACTT A TCATGA-3' 5'-AAGAAA C TCCACTT C TCATGA-3' 5'-AAGAAA G TCCACTT C TCATGA-3'	Ryan et al, <i>J. Natl. Cancer Inst.</i> 79, 1351 (1987)
Лекарственная зависимость, синдром Туретта, нарушение внимания и гиперактивность	<i>TDO2</i>	6 интрон 663 G/A 666 G/T	WT M1 M2	5'-AAATAATG G CA G ATAAGAAT-3' 5'-AAATAATG A CA G ATAAGAAT-3' 5'-AAATAATG G CA T ATAAGAAT-3'	Comings et al, <i>Pharmacogenetics</i> 6, 307 (1996)
Острая малярия	<i>NTF<math>\alpha</math></i>	Промотор -376 G/A	$\alpha$ G $\alpha$ A	5'-TCCTGTCTGGAA G TTAGAAGGA-3' 5'-TCCTGTCTGGAA A TTAGAAGGA-3'	Knight, et al, <i>Nature Gen.</i> 22, 145 (1999)
Дефицит белка С типа	<i>pC</i>	Промотор -114 C/T	WT MT	5'-TGTGGTTATGGA T TAACTCGAA-3' 5'-TGTGGTTATGGA C TAACTCGAA-3'	Spek et al, <i>J. Biol. Chem.</i> 273, 10168 (1998)
Синдром Бернарда - Соулье	<i>Gplb<math>\beta</math></i>	Промотор -133 C/G	WT MT	5'-TAGTGTGCTAT C TGCCGCTGCAG-3' 5'-TAGTGTGCTAT G TGCCGCTGCAG-3'	L. Ludlow et al, <i>J. Biol. Chem.</i> 271, 22076 (1996)
Гемофилия	<i>Factor VII</i>	Промотор -94 C/G	WT MT	5'-GTCCTCCCCTCC C CCATCCCTCT-3' 5'-GTCCTCCCCTCC G CCATCCCTCT-3'	Carew et al, <i>Blood</i> 92, 1639 (1998)
Фетальная форма гемоглобина	<i>G<math>\gamma</math> globin</i>	Промотор -114 C/T	WT MT	5'-CTTGCCTTGAC C AATAGCCTTG-3' 5'-CTTGCCTTGAC T AATAGCCTTG-3'	Fucharoen, et al, <i>Nucl. Acids Res.</i> 18, 5245 (1990)

Базы данных HGMD, dbSNP, HGBASE, ALFRED, OMIM содержат более 100 ОНП в регуляторных районах генов, а также более 10<sup>4</sup> ОНП в кодирующих частях и 10<sup>3</sup> ОНП в сайтах сплайсинга.



# Интернет-ресурс для анализа рОНП, rSNP\_Guide

В текущем релизе БД MATRIX представлено более 1500 последовательностей связывания 41 ТФ-сайта, экстрагированных из баз данных TRANSFAC и TRRD. Каждый TF-сайт представлен 8-16 олигонуклеотидными частотными матрицами.

## ШАГ 1: Предсказание неизвестного ТФ-сайта с последовательностью ДНК, содержащей рОНП

The screenshot displays the rSNP-Tools web interface. The main window shows a table for the prediction of TF-sites damaged/appeared due to SNPs. The table has columns for TF-site, SNP #1, SNP #2, SNP #3, and Euclidean distances. The TF-site column lists various transcription factors, with CP-1 highlighted in red. The SNP #1 and SNP #2 columns show values +0.3 and -0.5 respectively, also highlighted in red. The Euclidean distances column shows values X++, X+, X-, and X+.

Numbered callouts (1-7) indicate the workflow: 1. Selecting a TF-site (CP-1); 2. Selecting a SNP; 3. Inputting a DNA sequence; 4. Executing the prediction; 5. Viewing the prediction results; 6. Viewing the frequency matrix plot; 7. Viewing the frequency matrix plot.

The right window shows the CP-1 transcription factor binding DNA-site page. The input DNA sequence is G C A G A T A A G A A T A G G G A G. The Bases Available dropdown is set to SRSS from Heidelberg (EMBL) by ID. The Execute button is highlighted in red.

Two frequency matrix plots are shown below the main window. The left plot is titled 'CP-1: Mean Recognition - Frequency Matrix Score' and the right plot is titled 'CP-1: Mean Recognition - Frequency Matrix Score'. Both plots show a line graph of the frequency matrix score versus length in base pairs (bp).

Результат: каждый потенциальный ТФ-сайт характеризуется  $2 \times N$ -размерным численным вектором (N, число изученных вариантов ОНП).



# Экспериментальный анализ рОНП

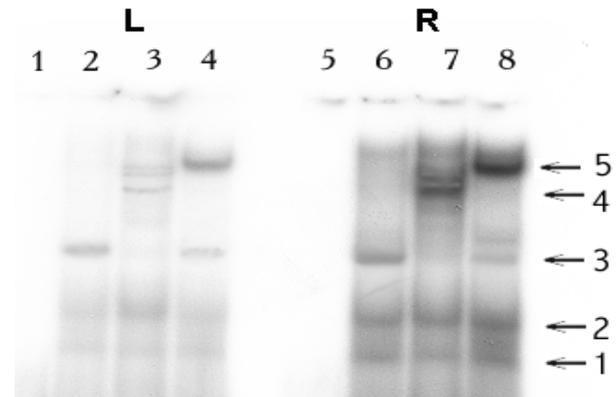
на примере ОНП в гене *TDO2* (Vasiliev G. *Et al.*, 1999, *FEBS L.*, 462, 85.)



1. Гель-шифт анализ тестирует ОНП по наличию комплексов с ядерными белками с различной подвижностью в геле.

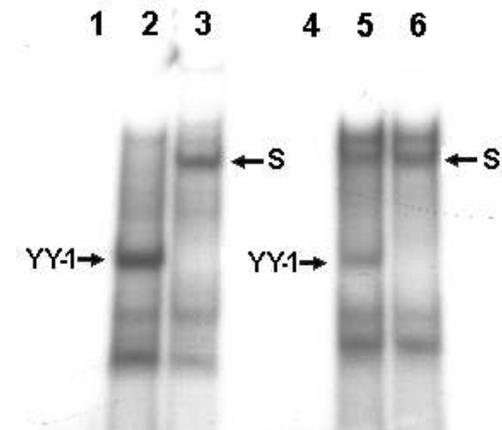
WT            5'-AAATAATG **G** CA **G** ATAAGAAT-3'  
M1            5'-AAATAATG **A** CA **G** ATAAGAAT-3'  
M2            5'-AAATAATG **G** CA **I** ATAAGAAT-3'

Связывание ядерных белков с олигонуклеотидами WT (2, 6), M1 (3, 7), M2 (4, 8); 1,5 - без яд. Экстракта. (L) - инкубация 1 час., (R) - 4 часа с ядерным экстрактом клеток печени крысы.



2. Поиск ТФ-кандидата, чей сайт связывания поврежден мутацией.

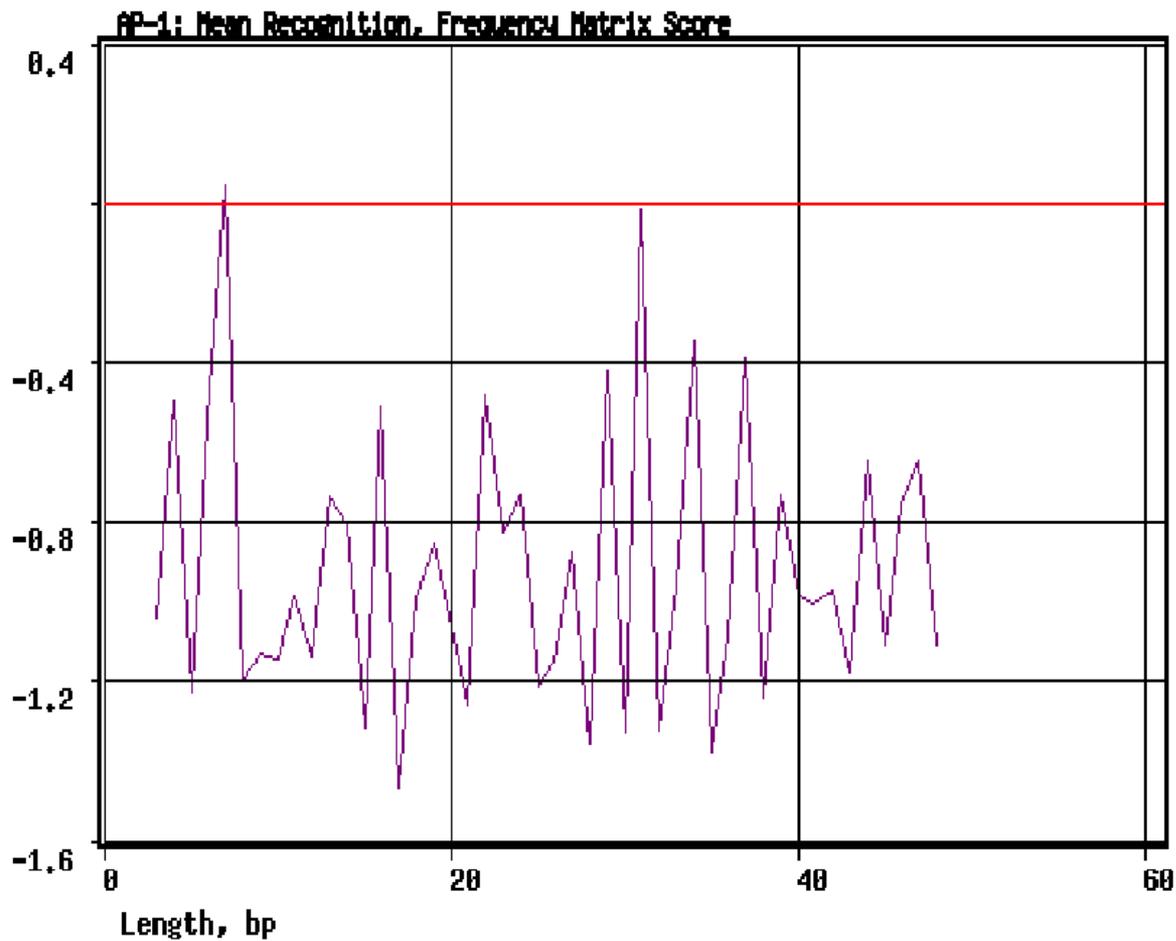
3. Экспериментальная проверка ТФ – кандидата (посредством связывания со специфическими ядерными экстрактами, антителами или олиго-компетиторами).



Действие анти-YY-1 антител на подвижность комплекса ядерных белков с WT (1, 2, 3) и M2 (4, 5, 6). Дорожки 1, 4, без экстракта. Дорожки 2, 5: 3  $\mu$ g ; Дорожки 3, 6: 3  $\mu$ g с преинкубацией со специфическими антителами к YY-1 фактору.



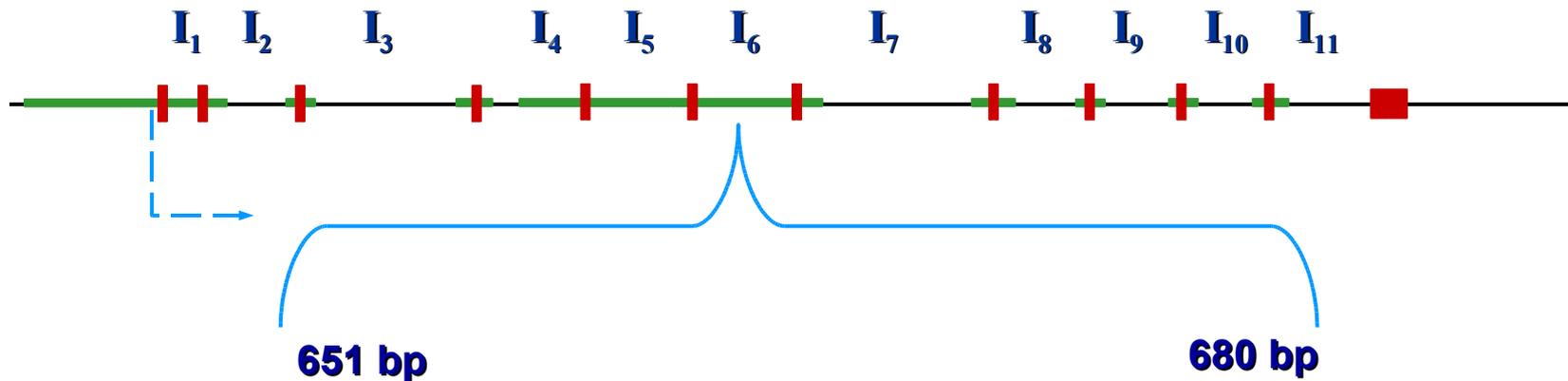
**B**





## Моделирование влияния мутаций:

мутации в 6-м интроне гена триптофан 2,3 диоксигеназы сопровождаются комплексом психических патологий, таких как рассеянное внимание, наркотическая зависимость, синдром Туретта и др



Норма : 5' -cagtTGCCAAATAATG **G** CAGATAAGAATAGGGAG - 3'

Мутация : 5' -cagtTGCCAAATAATG **A** CAGATAAGAATAGGGAG - 3'

**G** --> **A**



# Интернет-ресурсы для анализа рОМП



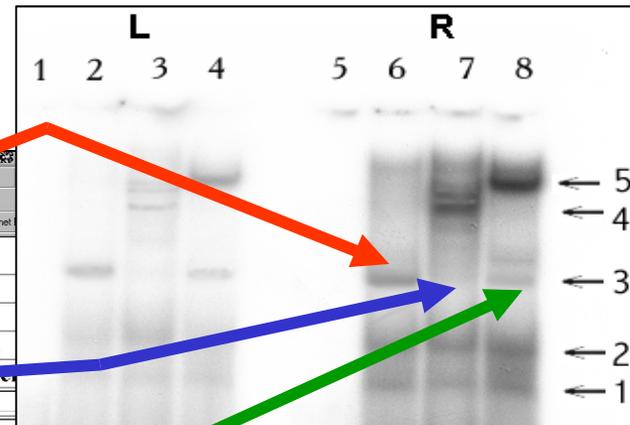
## ШАГ 2: ПОСТРОЕНИЕ 2×N-РАЗМЕРНЫХ ВЕКТОРОВ, X-САЙТОВ, СИМУЛИРУЮЩИХ ДАННЫЕ ГЕЛЬ-ШИФТ АНАЛИЗА

WT            5'-AAATAATG **G** CA **G** ATAAGAAT-3'  
 M1            5'-AAATAATG **A** CA **G** ATAAGAAT-3'  
 M2            5'-AAATAATG **G** CA **I** ATAAGAAT-3'

The screenshot shows the rSNP-Tools interface with the following data:

X-site	SNP #1	SNP #2	SNP #3	Euclidean distances	Prediction
I X++					
II X+-	+	0	+	0,5	
III X-+	+1,0	0	+0,5		
IV X--		-0,4			No idea

The 'Calculate' button is circled in red. Other buttons include 'Converter', 'Report', and 'Clear All'.



Связывание ядерных белков с олигонуклеотидами WT (2, 6), M1 (3, 7), M2 (4, 8); 1,5 - без ядерного экстракта. (L) - инкубация 1 час.;

(R) - инкубация 4 часа с ядерным экстрактом клеток печени.



rSNP-Tools - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Size Print Edit ReGet Messenger Links

### rSNP-Tools

#### Prediction of TF-sites damaged/appeared due to SNP

TF-site	SNP #1		SNP #2		SNP #3		Euclidean distances				Similarity			
	(+)	(-)	(+)	(-)	(+)	(-)	X++	X+-	X-+	X--	X++	X+-	X-+	X--
<input checked="" type="checkbox"/> <a href="#">MATRIX tools</a>														
<input checked="" type="checkbox"/> <a href="#">AP-1</a>	+0	-0.2												
<input type="checkbox"/> <a href="#">ATF</a>			+0.5											
<input type="checkbox"/> <a href="#">C/EBP</a>			+0.4											
<input type="checkbox"/> <a href="#">c-Fos</a>			+0.3											
<input type="checkbox"/> <a href="#">c-Jun</a>			+0.2											
<input type="checkbox"/> <a href="#">c-Jun</a>			+0.1											
<input type="checkbox"/> <a href="#">c-Jun</a>			+0											
<input type="checkbox"/> <a href="#">c-Myb</a>			-0.1											
<input type="checkbox"/> <a href="#">c-Myb</a>			-0.2											
<input type="checkbox"/> <a href="#">COUP</a>			-0.3											
<input type="checkbox"/> <a href="#">COUP</a>			-0.4											
<input type="checkbox"/> <a href="#">CP-1</a>			-0.5											
<input type="checkbox"/> <a href="#">CP-1</a>														
<input type="checkbox"/> <a href="#">CRE-BP1</a>														
<input type="checkbox"/> <a href="#">CREB</a>														
<input type="checkbox"/> <a href="#">E2</a>														
<input type="checkbox"/> <a href="#">E2F</a>														
<input type="checkbox"/> <a href="#">EN</a>														
<input type="checkbox"/> <a href="#">ER</a>														
<input type="checkbox"/> <a href="#">Ets</a>														
<input type="checkbox"/> <a href="#">GAGA</a>														
<input type="checkbox"/> <a href="#">GAGA</a>														
<input type="checkbox"/> <a href="#">GAL4</a>														
<input type="checkbox"/> <a href="#">GATA</a>														
<input type="checkbox"/> <a href="#">GR</a>														
<input type="checkbox"/> <a href="#">GR</a>														
<input type="checkbox"/> <a href="#">LIME1</a>														

My Computer

Start (D:\TEMP) - Far (P:\Julia\...rS... (D:\PROG\FV... (D:\PROG\FV... rSNP-Tools... Microsoft Pow... Microsoft Wor... rSNP\_Tools: P... 11:31 AM



Б

AP-1 transcription factor binding DNA-site - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Print Edit ReGet Messenger Links

## AP-1 transcription factor binding DNA-site

**Input DNA Sequence :**

**from Screen:**

GGCGCTGCGGCTCCGGCGAT&GTGTGCTAT G TGCCGCTGC&GCGCGCTCCGGCGGGCCTC

**from DB:**  **Bases Available:** SPS5 from Heidelberg (EMBL) by ID

**from File:**  Browse... [File formats here.](#)

Execute Reset form

---

**Select one of the AP-1 Frequency Matrices within the Oligonucleotide Alphabets listed below:**

- AP-1 within Alphabet {A, T, G, C}
- AP-1 within Alphabet {K, M}
- AP-1 within Alphabet {KK, KM, MK, MM}
- AP-1 within Alphabet {KKK, KKM, ..., MMK, MMM}
- AP-1 within Alphabet {R, Y}
- AP-1 within Alphabet {RR, RY, YR, YY}
- AP-1 within Alphabet {RRR, RRY, ..., YYR, YYY}
- AP-1 within Alphabet {W, S}
- AP-1 within Alphabet {WW, WS, SW, SS}
- AP-1 within Alphabet {WWW, WWS, ..., SSW, SSS}
- AP-1 within Alphabet {KxK, KxM, MxK, MxM}
- AP-1 within Alphabet {KxKxK, KxKxM, ..., MxMxK, MxMxM}

Start (D:\TEMP) (P:\Julia\... (D:\PROGF... (D:\PROGF... (SNP-Tools... Microsoft Po... Microsoft W... (SNP\_Tools... AP-1 tran... EN 11:37 AM



rSNP-Tools - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Size Print Edit ReGet Messenger Links

### rSNP-Tools

#### Prediction of TF-sites damaged/appeared due to SNP

TF-site	SNP #1		SNP #2		SNP #3		Euclidean distances				Similarity				
	(+)	(-)	(+)	(-)	(+)	(-)	X++	X+-	X-+	X--	X++	X+-	X-+	X--	
<input checked="" type="checkbox"/> <a href="#">MATRIX tools</a>															
<input checked="" type="checkbox"/> <a href="#">AP-1</a>	+0	-0.2													
<input type="checkbox"/> <a href="#">ATF</a>			+0.5												
<input type="checkbox"/> <a href="#">C/EBP</a>			+0.4												
<input type="checkbox"/> <a href="#">c-Fos</a>			+0.3												
<input type="checkbox"/> <a href="#">c-Jun</a>			+0.2												
<input type="checkbox"/> <a href="#">c-Jun</a>			+0.1												
<input type="checkbox"/> <a href="#">c-Jun</a>			+0												
<input type="checkbox"/> <a href="#">c-Myb</a>			-0.1												
<input type="checkbox"/> <a href="#">c-Myb</a>			-0.2												
<input type="checkbox"/> <a href="#">COUP</a>			-0.3												
<input type="checkbox"/> <a href="#">COUP</a>			-0.4												
<input type="checkbox"/> <a href="#">CP-1</a>			-0.5												
<input type="checkbox"/> <a href="#">CP-1</a>															
<input type="checkbox"/> <a href="#">CRE-BP1</a>															
<input type="checkbox"/> <a href="#">CREB</a>															
<input type="checkbox"/> <a href="#">CREB</a>															
<input type="checkbox"/> <a href="#">E2</a>															
<input type="checkbox"/> <a href="#">E2F</a>															
<input type="checkbox"/> <a href="#">E2F</a>															
<input type="checkbox"/> <a href="#">EN</a>															
<input type="checkbox"/> <a href="#">EN</a>															
<input type="checkbox"/> <a href="#">ER</a>															
<input type="checkbox"/> <a href="#">ER</a>															
<input type="checkbox"/> <a href="#">Ets</a>															
<input type="checkbox"/> <a href="#">Ets</a>															
<input type="checkbox"/> <a href="#">GAGA</a>															
<input type="checkbox"/> <a href="#">GAGA</a>															
<input type="checkbox"/> <a href="#">GAL4</a>															
<input type="checkbox"/> <a href="#">GAL4</a>															
<input type="checkbox"/> <a href="#">GATA</a>															
<input type="checkbox"/> <a href="#">GATA</a>															
<input type="checkbox"/> <a href="#">GR</a>															
<input type="checkbox"/> <a href="#">GR</a>															
<input type="checkbox"/> <a href="#">LTF1</a>															

My Computer

Start (D:\TEMP) - Far (P:\julia\...r... (D:\PROG\FV... (D:\PROG\FV... rSNP-Tools... Microsoft Pow... Microsoft Wor... rSNP\_Tools: P... EN 11:31 AM



Г

rSNP-Tools - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Refresh Home Search Favorites History Mail Size Print Edit ReGet Messenger Links

<input checked="" type="checkbox"/>	<u>USF</u>	-0.5	-0.2	+0	-0.1			1.92353	1.52565	1.26150	0.46801	-0.8943	-1.0633	-0.6875	0.73940										
<input checked="" type="checkbox"/>	<u>YY1</u>	+0.1	+0.3	+0.3	+0.1			1.18321	1.22654	1.05091	1.09947	-0.1540	-0.7642	-0.4769	0.10793										
<b>Select SNP's</b>		<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>				<b>Statistical analysis</b>				<b>Significance Level</b>													
<b>Average of Distribution</b>								1.73919	1.41881	1.23446	0.63561	0.005													
<b>Mean Square Deviation</b>								0.34230	0.34006	0.29883	0.44588	<b>Accuracy</b>													
								0.0																	
<b>Simulate the gel-shift data by assigning the X-site listed below</b>																									
<b>DNA-protein Binding</b>		<b>SNP #1</b>		<b>SNP #2</b>		<b>SNP #3</b>		<b>Euclidean distances</b>				<b>Prediction</b>													
		+1.0		0		-1.0																			
<b>X-site</b>		<b>(+)</b>		<b>(-)</b>		<b>(+)</b>		<b>(-)</b>		<b>X++</b>		<b>X+-</b>		<b>X-+</b>		<b>X--</b>		<b>TF-site is present</b>							
<b>I</b>	<b>X++</b>	1	1	0	0	-1	-1	0	1.38314	1.38314	1.38314	1.95606													
<b>II</b>	<b>X+-</b>	1	-0.3405	0	-0.3405	-1	-0.3405	1.38314	0	1.95606	1.38314														
<b>III</b>	<b>X-+</b>	-0.3405	1	-0.3405	0	-0.3405	-1	1.38314	1.95606	0	1.38314	GATA:													
<b>X-site simulates the absence of DNA-protein complex</b>																									
<b>IV</b>		<b>X--</b>		-0.3405		-0.3405		-0.3405		-0.3405		-0.3405		1.95606		1.38314		1.38314		0		AP-1; ATF; c-Fos; c-Jun; c-Myb			
<b>No idea</b>																									
<b>all (-)</b>		GR; MyoD; NF-E2; NF-IL6;																							
<b>Conflict</b>		RF-X								<b>Near zero</b>															

Calculate

Converter    Report    Clear All

EXAMPLES

My Computer

Start {D:\TEMP...} {P:\Julia\...} {D:\PRO...} {D:\PRO...} rSNP-T... Microsoft ... Microsoft ... rSNP\_Too... Find API 1... EN 12:12 PM



# Система rSNP\_Guide

**ШАГ 3:** Вычисление расстояния Евклида от симулированного X-сайта

до всех Z сайтов:  $D_{TF,X} = [\sum_{1 \leq n \leq 2 \times N} (\text{MAX\_Score}_n(X) - \text{MAX\_Score}_n(Z))^2]^{1/2}$

**ШАГ 4:** Проверка гипотезы  $\{D_{FT,X} < D_{FT,X-}\}$  на достоверность согласно критерию

Стьюдента,  $t_{\alpha,v}$ . Дает доказательство того, что X-сайт наиболее вероятно является известным ТФ-сайтом, чем характеризуется »отсутствием любого сайта:

$$d_{TF,X} = D_{X-,X} - D_{TF,X} \cdot t_{\alpha,v} \times \text{s.d.}(D_{TF,X}), \quad d_{TF,X-} = t_{\alpha,v} \times \text{s.d.}(D_{TF,X-}) - D_{TF,X-}$$

The screenshot shows the rSNP\_Guide web application interface. The main window displays a table with columns for 'Score', 'WT', 'M1', 'M2', 'Euclidean distances', and 'Student's t-scores'. The table lists various transcription factors (TFs) such as AP-1, ATF, c-Fos, c-Jun, c-Myb, COUP, CP-1, CRE-BP1, and CREB. The 'Score' column contains values for (+) and (-) variants. The 'Euclidean distances' column contains values for X and X- variants. The 'Student's t-scores' column contains values for X and X- variants. The table is filtered to show only TFs with a score greater than 0.05. The table is sorted by the 'Student's t-scores' column in descending order. The table is displayed in a grid format with a scroll bar on the right. The table is titled 'rSNP Tools: P0rSNP0J0004a' and 'SYSTEM: T0rSNP0J0004a'. The table is displayed in a browser window with the address bar showing 'http://www.mgs.bionet.nsc.ru/mgs/programs/rntp/images/p0rntp0j0004a.html'. The browser window title is 'rSNP Tools: P0rSNP0J0004a - Microsoft Internet Explorer'. The browser window also shows a search bar and a 'Ссылки' button. The browser window also shows a taskbar with other open applications like 'Microsoft Word - SNP\_T...', 'Microsoft PowerPoint - [r...', and 'Зона Интернета'. The taskbar also shows the system clock as '15:04'.

Score	WT	M1	M2	Euclidean distances	Student's t-scores
AP-1	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
ATF	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
c-Fos	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
c-Jun	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
c-Myb	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
COUP	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
CP-1	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
CRE-BP1	(+) (-)	(+) (-)	(+) (-)	X X-	X X-
CREB	(+) (-)	(+) (-)	(+) (-)	X X-	X X-

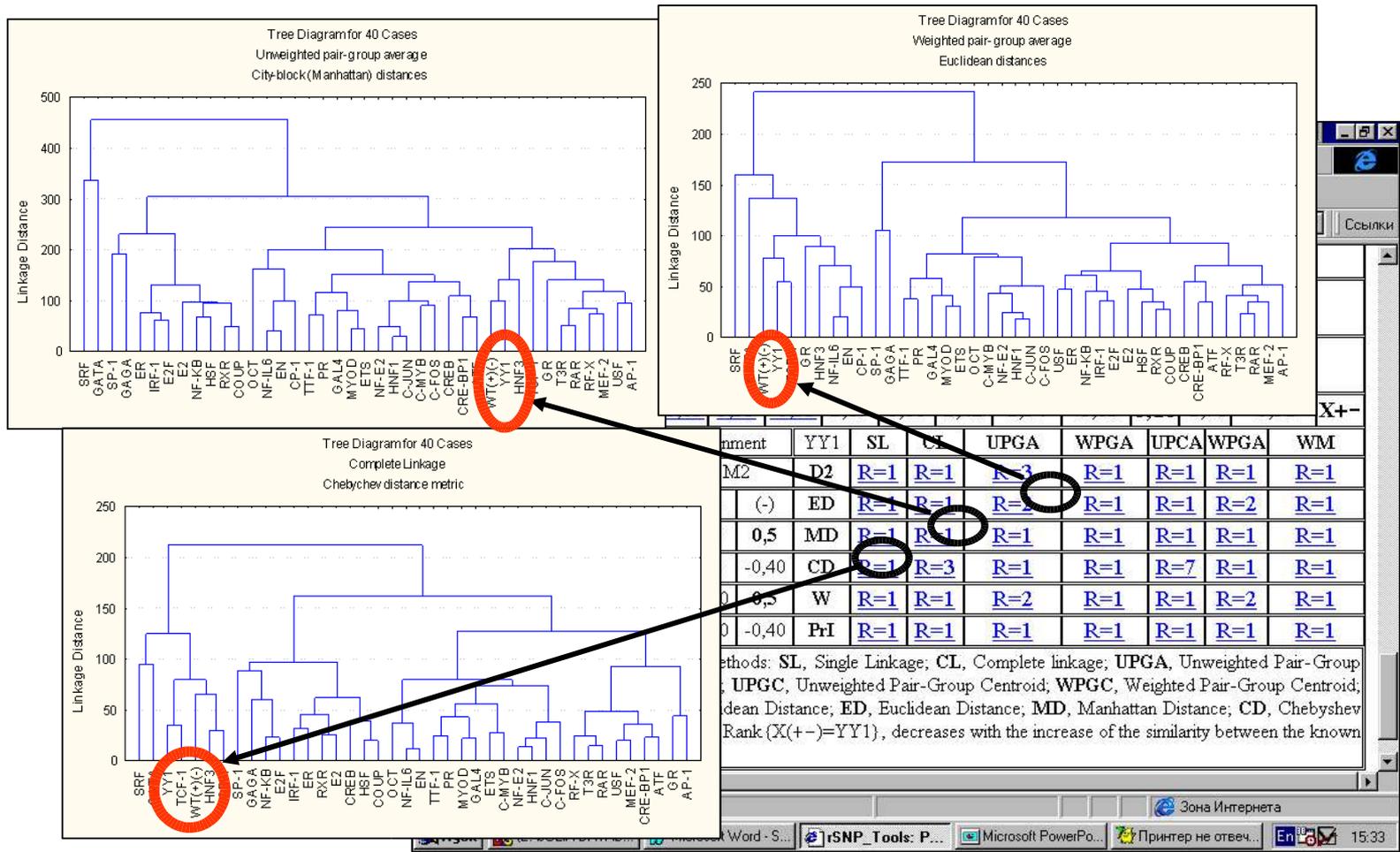
**ПРАВИЛО ПРЕДСКАЗАНИЯ:**  $\{d_{TF,X} > 0\} \& \{d_{TF,No} < 0\} \Rightarrow X \approx TF, \text{ ТФ-сайт представлен в виде X-сайта}$



# Система rSNP\_Guide



## ШАГ 4: проверка робастности (устойчивости) предсказания ТФ-сайта





# Результаты анализа рОИП системой rSNP\_Guide



SNP's sequence data		SNP's experimental data			rSNP_Tools analysis				Experimental control			
#	Gene	Name	Extract/cell	DNA/proteins			N-	N+	X-site	TF-site	Test	Reference
2	<i>TDO2</i>	WT M1 M2	Liver	++ - +-	- ++ -	+- - ++	168	74	WT+ M1+ M2+	YY1 GR HNF-3	anti-body none oligoDNA	Vasiliev et al, <i>FEBS Lett.</i> 462, 85 (1999)
1	<i>K-ras</i>	CA CC GC	Lung		+	- -	196	40	CA+	GATA	OligoDNA competition	Levashova et al., <i>BGRS'2000 Proceedings</i> (2000)
3	<i>NTF<math>\alpha</math></i>	$\alpha$ G(wt) $\alpha$ A(mt)	MonoMac6		- +		144	17	$\alpha$ A+	OCT	antibody	Knight, et al, <i>Nature Gen.</i> 22, 145 (1999)
4	<i>pC</i>	WT MT	HepG2		+	-	134	28	WT++	HNF1	antibody	Spek et al, <i>J. Biol. Chem.</i> 273, 10168 (1998)
5	<i>Gplb<math>\beta</math></i>	WT MT	CHRF-288		+	-	111	45	WT+	GATA	antibody	Ludlow et al, <i>J. Biol. Chem.</i> 271, 22076 (1996)
	<i>Factor VII</i>	WT MT	HepG2		+	-	107	52	WT+	Sp1	antibody	Carew et al, <i>Blood</i> 92, 1639 (1998)
	<i><math>\gamma</math>globin</i>	WT MT	MEL HeLa		+	-	104	52	WT+	CP-1	OligoDNA competition	Fucharoen, et al, <i>Nucl. Acids Res.</i> 18, 5245 (1990)

Недостаток системы: число **положительных** скоров распознавания ТФ-сайтов очень велико (в пределах 10- 30% от общего количества)